

Habilitation à Diriger des Recherches
Mathématiques



MARCHES ET DIFFUSIONS, MILIEUX ALÉATOIRES & APPLICATIONS

Pierre Andreoletti

Rapporteurs :

| | |
|----------------|--------------------------------|
| Francis Comets | Université Paris Diderot |
| Nina Gantert | Technische Universität München |
| Yueyun Hu | Université Paris 13 |

Soutenue le 28 Novembre 2014 devant le jury composé de :

| | |
|--------------------|--|
| Romain Abraham | Université d'Orléans, Président |
| Francis Comets | Université Paris Diderot |
| Nathanaël Enriquez | Université Paris Ouest Nanterre La Défense |
| Yueyun Hu | Université Paris 13 |
| Pierre Picco | CNRS, I2M - Marseille |
| Zhan Shi | Université Pierre et Marie Curie |

REMERCIEMENTS

Je remercie Francis Comets, Nina Gantert et Yueyun Hu qui ont accepté de faire un rapport sur mon travail. Je leur suis très reconnaissant de m'avoir accordé de leur temps. Une mention particulière à Francis Comets qui a déjà rapporté sur ma thèse de doctorat et qui a depuis un regard bienveillant sur mon travail.

Zhan Shi apparaît lui aussi pour la seconde fois dans mon jury, ses travaux ont guidé bon nombre des miens, c'est donc un honneur pour moi qu'il soit à nouveau présent. Merci aussi à Nathanaël Enriquez d'avoir accepté de faire partie de ce jury. Ses articles, source d'inspiration, m'accompagnent souvent ces derniers temps.

Je remercie Pierre Picco qui m'a initié à ce beau métier où l'on oscille entre doute et certitude - parfois déroutant, parfois enivrant. Parmi ses meilleurs conseils, je retiendrai celui de la recherche du meilleur scénario, pour les probas bien sûr, mais dans la vie aussi !

Enfin, Romain Abraham m'a fait confiance dès mon arrivée à Orléans en me proposant, notamment, un co-encadrement de thèse. Cela a été déterminant pour la suite, je lui dois donc beaucoup.

Je remercie Pierre (D) et Roland, ensemble nous avons fait un bout de chemin - la route est encore (agréablement) longue, Alexis et sa quête de l'ultime détail mathématique, Catherine et Dasha avec qui je partage de biens jolis problèmes appliqués. Plus récemment j'échange avec Grégoire mais aussi avec Bastien et Xinxin tous incroyablement enthousiastes, le regard neuf. Enfin grâce à Maxime et Pierre (G) je me suis essayé à un peu d'économie, pourquoi pas !

Je remercie également les membres dynamiques du MAPMO, ceux avec qui je discute régulièrement, de math ou autre, les collègues du département informatique de l'IUT dont l'ouverture d'esprit m'a permis de concilier au mieux enseignement et recherche. Je remercie aussi l'équipe qui nous soutient au quotidien, Anne, Marie-France, Marie-Laurence et bien entendu Romain (T).

J'exprime ma reconnaissance (et au delà) à Laurence qui observe, paisiblement, les différentes phases de la recherche en mathématiques - m'interrogeant parfois : mais est-ce bien sérieux tout ça ? Elle contribue largement à mon équilibre.

En réalisant ce travail, je pense à Camille & Clément dont les échanges passés continuent à orienter mes choix de vie, à Oriana, Mylann & Ambre, ainsi qu'à ma famille et mes amis. Ils contribuent chacun à leur manière à l'épanouissement d'une vie.

Table des matières

| | |
|---|-----------|
| Introduction | 5 |
| 1 Temps et espace discrets | 9 |
| 1.1 Localisation du cas récurrent | 9 |
| 1.2 Concentration | 10 |
| 1.3 Prolongements | 12 |
| 2 Temps et espace continus | 14 |
| 2.1 Le temps local du cas récurrent | 14 |
| 2.2 Localisation du cas transient lent | 17 |
| 2.3 Le temps local du cas transient lent | 20 |
| 2.4 Prolongements | 23 |
| 3 Temps discret, espace branchant discret | 24 |
| 3.1 Les travaux de Y. Hu et Z. Shi | 25 |
| 3.2 La plus grande génération entièrement visitée | 26 |
| 3.3 Nombre de sites visités à une génération donnée | 28 |
| 3.4 Etalement des points visités | 30 |
| 3.5 Prolongements | 33 |
| 4 Applications | 36 |
| 4.1 Estimation pour la marche de Sinai | 36 |
| 4.2 Dégrafage d'ADN | 37 |
| 4.3 Estimation en milieu Markovien - Lien entre MAMA et CMC | 41 |
| 4.4 Prolongements | 43 |

Introduction

Les thèmes abordés dans cette synthèse traitent de processus en milieu aléatoire. Introduits à la fin des années 60 ils traduisent la complexité du mouvement d'une particule soumis à plusieurs aléas. Ces aléas provenant d'une part du caractère aléatoire propre à la particule et d'autre part de la complexité ou de la connaissance partielle du substrat sur lequel se déplace la particule.

La richesse du modèle vient du fait qu'il ne se limite pas à un unique type de processus mais plutôt à un ensemble de processus avec multiples niveaux de stochasticité. Ainsi, à son introduction, il est aussi bien utilisé par le bio-physicien A. A. Chernov [Che62], pour un modèle de réplication de l'ADN que par des physiciens D. E. Temkin [Tem69], [Tem72] pour la croissance de cristaux, puis plus tard par Ya. G. Sinai [Sin82b] pour la description d'un gaz de Lorentz.

Nous verrons que chaque type de milieu aléatoire que nous aborderons ici aura sa spécificité propre, ceci nous mènera à la manipulation de différents autres processus aléatoires plus élémentaires.

Ce travail est organisé en 4 parties, la première partie traite du cas le plus élémentaire de Marches Aléatoires en Milieu Aléatoire, il s'agit de marches discrètes unidimensionnelles au plus proche voisin évoluant sur un potentiel aléatoire, une partie de ces travaux provenant de ma thèse. La seconde partie a pour objet le cas équivalent mais en temps et espace continus, plus particulièrement il s'agira de diffusion Brownienne sur potentiel Brownien. La troisième partie traite d'une marche discrète évoluant sur un arbre Galton-Watson, l'environnement étant alors donné par une marche branchante. Enfin la dernière partie est une ouverture des résultats théoriques à des problèmes d'estimation.

Partie 1 : Les MAMA sont étudiées par les mathématiciens depuis les années 70, cependant cette thématique a connu un essor important au début des années 2000 et continue à l'être aujourd'hui. Les articles fondateurs sont ceux de F. Solomon [Sol75] (1975) qui classe ces modèles en terme de critère de récurrence, ceux de H. Kesten M.V. Kozlov F. Spitzer [KKS75] (1975) qui affine cette classification dans le cas transient en étudiant leur comportement en temps long, et enfin Sinai [Sin82a] (1982) qui résout le cas récurrent.

Je reviendrai sur la marche de Kesten-Kozlov-Spitzer dans la Partie 2 lorsque je parlerai du cas transient en temps et espace continus qui présente de nombreuses similarités avec son homologue discret. Je m'intéresserai donc dans ce paragraphe uniquement au cas récurrent. La marche récurrente, traditionnellement appelée marche de Sinai, doit son nom au travail de Ya. G. Sinai [Sin82a] qui en explique son comportement asymptotique. L'idée intuitive qu'il donne à propos de son résultat est la suivante « The moving point accumulates the information about the realization of the environment and then gets stuck in regions of the strongest fluctuations of the environment » [Sin82b]. Ainsi comprendre quel est le mouvement typique de la marche à un instant donné n revient à comprendre où est-ce que l'environnement effectue sa plus grande fluctuation en lien étroit avec n . La représentation que l'on doit avoir à l'esprit ici est une marche aléatoire attirée par les puits de potentiels créés par l'environnement. C'est à cet aspect d'emprisonnement que je me suis intéressé en donnant dans un premier temps une preuve alternative à la localisation de Sinai [1]. Une question naturelle était aussi de savoir comment cette marche occupe son temps en regardant toute la trajectoire (jusqu'au temps n). J'ai alors montré, dans un premier temps par une convergence en probabilité dans [2] puis par la suite une convergence presque sûre dans [3] et [5], que la marche peut dépenser la quasi totalité de son temps n dans un voisinage de taille fini du réseau.

Partie 2 : Un article de Z. Shi s'intitule "Local time curiosity in random environment" [Shi98], il traite du temps local d'une diffusion Brownienne en environnement Brownien, c'est ce travail qui a éveillé ma curiosité pour le cas continu. Le modèle est l'équivalent continu de la MAMA dont l'origine remonte au travail de thèse de Schumacher [Sch84], [Sch85]. Le milieu Brownien considéré comportera une dérive (négative ou nulle) que je noterai $-\kappa/2$.

Lorsque cette dérive est nulle, la diffusion a le même comportement que la marche de Sinai comme le montre T. Brox [Bro86], puis Y. Hu et Z. Shi [HS98] pour des convergences presque sûres. Cependant il se passe quelque chose au niveau du temps local. Pour avoir un point de comparaison, revenons à la marche aléatoire simple dont les incréments indépendants sont de plus ou moins une unité avec probabilité $1/2$. Un travail de P. Révész (voir [R89] - chapitre 10), montre qu'il existe un principe d'invariance entre le temps local de cette marche usuelle et celui du processus de Wiener, c'est un résultat très fort puisque l'ensemble des temps locaux sont comparés sur un intervalle de temps. On pourrait donc s'attendre à ce que cela soit aussi le cas si on compare le temps local de la marche de Sinai et celui de la diffusion de Brox. Il n'en est rien : alors que la limite en loi pour le supremum des temps locaux des deux processus est semblable (voir N. Gantert Y. Peres et Z. Shi [GPS10] et P.A. et R. Diel [11]), cela ne se passe pas de la même façon quand on regarde les limites supérieures presque sûres. Alors que l'un borné par le temps imparti n converge lorsqu'on le divise par n ([R88], [GPS10]), le second normalisé par son temps t diverge, d'où cette curiosité remarquée par Shi. Remarquons que la différence de nature entre les temps locaux du cas discret et du cas continu n'est pas suffisante à expliquer ce phénomène, car sans milieu aléatoire cet aspect n'intervient pas. Le problème de la bonne normalisation des limites supérieure et inférieure pour le supremum des temps locaux de la diffusion de Brox a finalement été résolu par R. Diel [Diel1a].

Pour ce modèle continu je me suis également intéressé au cas avec dérive faible ($0 < \kappa < 1$) avec A. Devulder puis G. Vechambre. Dans ce cas la diffusion est transiente à droite mais sous-ballistique. Les travaux fondateurs pour cet équivalent continu de la marche discrète de Kesten-Koslov-Spitzer sont ceux de K. Kawazu H. Tanaka [KT97]. Nous avons eu une approche "quenched" c'est à dire comprendre le processus à environnement fixé pour des environnements fortement probables. Cette démarche avait été faite avant nous dans le cas discret par N. Enriquez, C. Sabot, O. Zindy [ESZ09c], [ESZ09b], [ESZ09a]. Le fait remarquable sous cette hypothèse de milieu faiblement atténué par la dérive, est l'apparition de grands puits de potentiel, suffisamment profonds pour retenir la diffusion mais jamais assez pour qu'au final la marche ne voit qu'un seul puits comme dans le cas de la diffusion de Brox. Ce phénomène de méta-stabilité associé à la transience de la marche fait apparaître une structure de renouvellement qui est à la base du comportement du processus.

Dans un premier travail [10] nous avons ainsi étendu les travaux de N. Enriquez C. Sabot et O. Zindy [ESZ09a] à cette diffusion, en montrant entre autre les phénomènes de localisation et de vieillissement. Dans un second travail [15], nous avons exploité plus en profondeur la structure de renouvellement pour étudier le temps local. L'étude jointe de la somme des temps de sortie des puits et des temps locaux dans le fond de ces puits permet, par exemple, d'explicitier la loi limite du supremum des temps locaux normalisé par t . La méthode basée sur la preuve d'un théorème fonctionnel ouvre la voie à une étude plus approfondie en vue d'obtenir des convergences presque-sûres.

Partie 3 : S'affranchir des MAMA uni-dimensionnelle n'est pas chose aisée. En particulier les difficultés qui apparaissent sur \mathbb{Z}^d , ne serait-ce que pour obtenir une classification grossière de ces MAMA restent encore majoritairement un challenge. De nombreuses avancées sur différents aspects ont cependant donné lieu à de nombreux travaux dont je ne parlerai pas ici, voici cependant deux références générales [Zei01], [DR13]. Un bon compromis est alors, peut-être, le

cas d'une MAMA dont l'espace à parcourir est un arbre, c'est en tout cas de celle-ci dont il sera question dans cette partie. Les résultats sur ce type de marche remontent aux années 90 où dans un article fondateur R. Lyons et R. Pemantle [LP92] déterminent des critères de récurrence dans le cas d'un arbre régulier. Ces résultats sont étendus par M. V. Menshikov et D. Petritis [MP02] puis par G. Faraud [Far11] pour des arbres de Galton-Watson. Comme pour le cas unidimensionnel, le comportement asymptotique de (X_n, n) en temps long dépend fortement des caractéristiques du milieu qui sont résumées dans une transformée de log-Laplace ψ . Je présenterai dans cette synthèse uniquement les cas récurrents.

Il se trouve que les marches récurrentes présentent une multitude de comportement asymptotique, les cas récurrents peuvent donc eux même être classifiés. Il s'agit avant tout de savoir à quel type d'objet on s'intéresse, Y. Hu et Z. Shi dans leurs premiers travaux ([HS07a], [HS07b]) étudient la plus grande génération visitée avant l'instant n que nous noterons comme eux X_n^* . Ils trouvent 3 comportements asymptotiques distincts, il s'agit de résultats presque sûrs sur l'ensemble de non extinction de l'arbre. Un cas extrêmement lent où $X_n^* \sim \log n$, un cas rapide où $\log X_n^* / \log n \sim 1 - 1/\min\{\kappa, 2\}$ ($\kappa := \inf\{s > 1, \psi(s) = 0\}$) et un cas intermédiaire sous-diffusif où $X_n^* \sim (\log n)^3$. Ce dernier cas concerne en fait deux marches récurrentes dont l'une est récurrente positive et l'autre récurrente nulle. G. Faraud Y. Hu et Z. Shi [FHS12], précisent par la suite la constante obtenue après re-normalisation. Leurs résultats et les techniques utilisés ouvrent alors la voie à d'autres travaux.

On a, par exemple, envie de décrire le nuage de points visités par la marche jusqu'à un instant donné n , autrement dit $\mathcal{V} := \{x \in \text{Arbre}, \mathcal{L}(x, n) \geq 1\}$ où \mathcal{L} désigne le temps local. Le but étant d'avoir un résultat quantitatif, par exemple la taille de \mathcal{V} , et également de pouvoir caractériser les sites de \mathcal{V} . Avec P. Debs [8] je me suis intéressé au problème de la plus grande génération entièrement visitée avant l'instant n , je la noterai R_n . Nous avons obtenu la convergence presque sûre de R_n qui se comporte dans tous les cas en $\log n$. En fait il y a une différence entre les cas récurrents lents et les autres du fait de la constante limite.

La question suivante est de savoir ce qu'il se passe entre R_n et X_n^* , en particulier prenons ℓ déterministe entre ces deux variables. Quel est le nombre typique de points visités à cette génération ℓ ? Je me limiterai au cas intermédiaire récurrent nul et supposerai donc $\ell = (\log n)^{1+\zeta}$ avec $0 < \zeta < 2$. Nous avons étudié avec P. Debs [9] le nombre de sites distincts de la génération ℓ visités avant n retour en la racine de l'arbre : $K_n(\ell)$. La moyenne annealed de $K_n(\ell)$ présente une transition de phase en $\zeta = 1$. Plus précisément, pour n grand fixé, le nombre de sites visités par génération augmente jusqu'à la génération $(\log n)^2$ puis diminue jusqu'à la génération $(\log n)^3$. Sur l'aspect occupation de l'espace, c'est à dire la façon dont la particule diffuse sur l'arbre, nous obtenons des résultats moins précis, mais qui donnent une idée de la diffusion de la marche sur l'arbre. En fait deux phénomènes antagonistes co-existent, d'une part on peut trouver régulièrement sur l'arbre des ensembles de sites visités très éloignés les uns des autres, et d'autre part de gros clusters de points contigus visités.

Partie 4 : Dans cette partie apparaissent des idées connexes aux MAMA unidimensionnelles, il n'est pas question d'obtenir de nouveaux résultats théoriques, mais plutôt d'adapter la démarche formelle de la Partie 1 à des problèmes appliqués. Pour l'essentiel il s'agit de problèmes inverses : en supposant qu'un processus noté $(X_n, n \in \mathbb{N})$ est guidé par un milieu donné ω sous-jacent, quelle information nous apporte la ou les trajectoires de $(X_n, n \in \mathbb{N})$ sur ω ?

L'étude de la concentration rencontrée dans la Partie 1, implique l'étude du temps local dans le voisinage du puits de potentiel, en étendant le plus possible ce voisinage et en repérant la coordonnée du fond du puits par l'intermédiaire du point favori on aboutit au travail [7]. On montre que pour la marche de Sinai, le logarithme du temps local peut être utilisé comme estimateur des différences du potentiel.

Un autre point de vue est utilisé par les bio-physiciens U. Bockelmann, B. Essevaz-Roulet, et F. Heslot. ([BERH97], [BERH98]) et physiciens V. Baldazzi, S. Cocco, E. Marinari, et R. Monasson ([BCMM07]) pour le séquençage d'une molécule d'ADN. L'idée est que la suite d'énergies (le milieu aléatoire) liant entre eux les deux brins de la molécule est caractérisée par la suite des bases complémentaires la constituant. En exerçant une force constante sur les deux brins de la molécule celle-ci peut être dégrafée, cependant la molécule a tendance à se reformer ce qui entraîne un mouvement de va et vient semblable à une marche aléatoire. La marche observée X représente alors le nombre de bases dégrafées à chaque instant jusqu'à ce que la molécule soit entièrement ouverte.

Avec R. Diel ([12]) nous avons repris cette étude dans le but de formaliser et améliorer les résultats obtenus par les physiciens, en particulier sur l'Estimateur du Maximum de Vraisemblance de la suite de bases. Remarquons que l'hypothèse d'indépendance du milieu aléatoire assez classique pour l'étude des MAMA ne peut être utilisée ici car il existe une dépendance entre les énergies de liaison.

Notons que dans les deux exemples précédents il n'est pas question d'estimer la distribution du milieu aléatoire mais plutôt ses trajectoires. Une première approche de la reconstruction de la loi du milieu à partir d'une trajectoire de la marche est donnée par O. Adelman et N. Enriquez [AE04]. Une seconde approche, dans le cas de marches unidimensionnelles, a été introduite par F. Comets, M. Falconnet, A. Gloter, D. Loukhianova et C. Matias, dans une série de 4 articles ([CFL⁺14], [FLM14],[MAL],[CFLL]) ils étudient dans les cas transient et récurrent l'EMV des paramètres de la distribution du milieu aléatoire. Avec Dasha Loukhianova et Catherine Matias ([13]), nous nous sommes intéressés au cas où le milieu n'est plus i.i.d. mais Markovien. Une des motivations étant de pouvoir traiter le dégrafage de l'ADN dont on sait que les probabilités de transition ne sont pas i.i.d. Une première étude de l'asymptotique de l'EMV a été facilitée par des résultats de R. Douc, É. Moulines et T. Rydén [DMR04] sur des chaînes de Markov cachées généralisées.

1 Temps et espace discrets

Une marche aléatoire en milieu aléatoire au plus proche voisin sur \mathbb{Z} , $(X_n, n \in \mathbb{N}; X_0 = 0)$, est l'association d'une suite de variables aléatoires $\omega := (\omega_x, x \in \mathbb{Z})$ à valeurs dans $(0, 1)^{\mathbb{Z}}$ et d'une chaîne de Markov à valeurs dans \mathbb{Z} dont les probabilités de transition, pour une suite fixée ω , sont données par $p(x, x+1) = \omega_x$ et $p(x, x-1) = 1 - \omega_x$. On notera ainsi \mathbb{P}^ω la mesure de probabilité (quenched) associée à cette chaîne de Markov, et \mathbb{P} la mesure de probabilité (annealed) induite par les deux objets et qui désignera donc la mesure de $(X_n, n \in \mathbb{N})$. \mathbb{P}_a ou \mathbb{P}_a^ω désigneront les mesures de probabilité conditionnées par l'événement $\{X_0 = a\}$. Les espérances relatives seront notées respectivement \mathbb{E}^ω , \mathbb{E} , \mathbb{E}_a et \mathbb{E}_a^ω .

Le comportement local de la marche s'interprète comme un processus discret sur un potentiel $\mathbf{S} := (S_x, x \in \mathbb{Z}, S_0 = 0)$, défini par $S_x := \sum_{i=1}^x \log \tilde{\omega}_i$ si $x > 0$ et $S_x := -\sum_{i=x+1}^0 \log \tilde{\omega}_i$ si $x < 0$ avec $\tilde{\omega}_i := (1 - \omega_i)/\omega_i$. Les probabilités de transition se ré-écrivent alors $\omega_x = e^{S_{x-1} - S_x} / (1 + e^{S_{x-1} - S_x})$ et on voit, avec cette écriture, que X a tendance à s'orienter vers les minima locaux de \mathbf{S} . Ce comportement local donne la trame de ce que sera le comportement en temps long de X .

Dans cette partie on va supposer que l'environnement ω est composé de variables aléatoires indépendantes et de même loi. Solomon [Sol75] donne une première classification de X en terme de critère de récurrence : $(X_n, n \in \mathbb{N})$ est récurrente si et seulement si $\mathbb{E}(S_1) = 0$.

1.1 Localisation du cas récurrent

On est ici dans le cas où les fluctuations de \mathbf{S} croissent ou décroissent avec la même amplitude, en particulier, dans un voisinage m de l'origine, un puits de potentiel de hauteur \sqrt{m} est créé avec une grande probabilité. Ceci est la clé de l'approche de Sinai [Sin82a]. L'énoncé de son résultat qui traduit le comportement de X en temps long nécessite la notion de *h-vallée* que l'on définit ici à partir d'un *h-minima* [introduit par J. Neveu J. Pitman [NP89] dans le cas continu]. Pour $h > 0$, on dit que $x \in \mathbb{Z}$ est un *h-minimum* pour \mathbf{S} s'il existe $u < x < v$ tel que $S_y \geq S_x$ pour tout $y \in \llbracket u, v \rrbracket$, $S_u \geq S_x + h$ et $S_v \geq S_x + h$. De plus x est un *h-maximum* pour \mathbf{S} si x est un *h-minimum* pour $-\mathbf{S}$. Une *h-vallée* est un intervalle $\llbracket u, v \rrbracket$ contenant un *h-minima* x réalisant le minimum de \mathbf{S} dans $\llbracket u, v \rrbracket$, tel que $S_u = \max_{z \in \llbracket u, x \rrbracket} S_z$ et $S_v = \max_{z \in \llbracket x, v \rrbracket} S_z$.

Notons maintenant m_h le *h-minima* le plus proche de la coordonnée zero du processus $(S_n, n \in \mathbb{Z})$. Aussi, sauf mention contraire, nous travaillerons, en plus de l'hypothèse de récurrence, sous l'hypothèse d'uniforme ellipticité : il existe $A > 0$ tel que $P(|S_1| \leq A) = 1$. Sinai donne le résultat de localisation suivant

Théoreme 1.1 ([Sin82a]).

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow +\infty} \mathbb{P} [|X_n - m_{\log n}| \leq \delta \log^2 n] = 1.$$

Ce résultat montre essentiellement deux choses. La première est que le comportement typique de X_n est en $(\log n)^2$: en un temps n , X ne pourra jamais franchir une barrière de potentiel générée par \mathbf{S} plus grande que $(1 + \epsilon) \log n$, mais franchira toute barrière plus petite que $(1 - \epsilon) \log n$. Or par hypothèse les premières fluctuations de \mathbf{S} de l'ordre de $\log n$ sont à une distance $(\log n)^2$ de l'origine.

La seconde est la notion de localisation locale dans le temps : pour n grand, X_n est dans le voisinage de $m_{\log n}$. Je donne ci-dessous très sommairement l'idée de la preuve donnée dans [1] et une autre approche qui sera utilisée dans la Section 2 :

Idées de preuve

• Dans le premier article de ma thèse [1], en utilisant l'existence de moments exponentiels pour S_1 plutôt que l'hypothèse d'uniforme ellipticité, je montre que pour tout instant plus petit que $n - e^{(\log n)^{3/4+\epsilon}}$ ($\forall \epsilon > 0$), la marche retournera en $m_{\log n}$ avec une grande probabilité. Ceci se fait en contrôlant les premier et second moments du minimum entre un temps de retour en $m_{\log n}$ pour X et le temps d'atteinte de $M_{\log n}^\pm$ qui est la coordonnée la plus proche (à droite avec le $^+$ et à gauche avec le $^-$) de $m_{\log n}$ tel que $S_{M_{\log n}^\pm} - S_{m_{\log n}} > \log n$. Notons que l'ensemble des points $W_n := \llbracket M_{\log n}^-, M_{\log n}^+ \rrbracket$ forme une $\log n$ -vallée de minima $m_{\log n}$. Une fois ceci obtenu nous n'avons plus qu'à refaire le même raisonnement que précédemment, à savoir que partant de $m_{\log n}$ en un temps restant $q_n := e^{(\log n)^{3/4+\epsilon}}$, X ne peut s'éloigner qu'au plus de $(\log q_n)^2 = (\log n)^{3/2+2\epsilon}$ de ce point de départ qui est négligeable devant $(\log n)^2$. On obtient donc une taille de voisinage de localisation légèrement améliorée par rapport au résultat original sans cependant parvenir au résultat de A. O. Golosov [Gol84] ni même celui obtenu plus tard par A. Bovier et A. Faggionato [BF08].

• La seconde approche vient du fait que l'on peut coupler localement dans le temps X et une seconde marche \tilde{X} qui est conditionnée à rester dans l'intervalle W_n . On sait que $\tilde{\mathbb{P}}_{m_{\log n}}^\omega(\tilde{X}_n \in \llbracket m_{\log n} \pm \delta \log n^2 \rrbracket) = \mu_n(\llbracket m_{\log n} \pm \delta \log n^2 \rrbracket)$, où μ_n est la mesure de probabilité invariante pour \tilde{X} , $\llbracket x \pm a \rrbracket$ est l'ensemble d'entiers $\llbracket x-a, x+a \rrbracket$ et $\tilde{\mathbb{P}}$ la distribution de \tilde{X} . En utilisant les (bonnes) propriétés de l'environnement \mathbf{S} recentré en $m_{\log n}$ on montre que $\mu_n(\llbracket m_{\log n} \pm \delta \log n^2 \rrbracket)$ tend vers 1 en probabilité. C'est cette approche qu'utilise Zeitouni dans [Zei01], Brox [Bro86] dans l'équivalent continu de la marche de Sinai et que nous avons utilisé dans le cas de la diffusion transiente lente [10]. \square

\triangleright Dans la preuve de Sinai [Sin82a], l'étude des excursions de la marche en dehors de $m_{\log n}$ est utilisée, cela nous amène naturellement à étudier les temps locaux dans la $\log n$ -vallée entre deux retours en $m_{\log n}$. On s'aperçoit alors que la marche dépense énormément de temps dans le fond de cette vallée, le paragraphe suivant étudie donc cet aspect. \triangleright

1.2 Concentration

Dans cette section et les suivantes il sera donc question de temps local \mathcal{L} en un point $x \in \mathbb{Z}$ ou sur un ensemble $A \subset \mathbb{Z}$ jusqu'à un instant $T \in \mathbb{N}^*$:

$$\mathcal{L}(T, x) := \sum_{i=1}^T \mathbf{1}_{X_i=x}, \quad \mathcal{L}(T, A) := \sum_{x \in A} \mathcal{L}(T, x)$$

Notons (u_n, n) une suite croissante d'entiers positifs telle que $\lim_{n \rightarrow +\infty} u_n = +\infty$. Le premier résultat de [2] (qui fait parti de mon travail de thèse) peut être résumé sous la forme

Théoreme 1.2 (A. [2]).

$$\lim_{n \rightarrow +\infty} \mathbb{P}(\mathcal{L}(n, \llbracket m_{\log n} \pm u_n \rrbracket)/n \geq 1 - u_n^{-1}) = 1$$

Ainsi en un temps n la marche de Sinai dépense la quasi totalité de son temps dans un petit voisinage (en comparaison des fluctuations typiques de X) de $m_{\log n}$. Cela n'empêche pas, bien entendu, à la marche de visiter des hauteurs de potentiels importantes mais ces temps excursions auront une contribution (en temps) négligeable, j'ai donc appelé ce phénomène concentration. Ce résultat conduit naturellement au comportement du temps local dans le voisinage de $m_{\log n}$. Notons $T_x := \inf\{k > 0, X_k = x\}$ le temps d'atteinte de x par X , on a

Proposition 1.3 (A. [2]). *Soit $n > 0$ il existe \mathcal{E}_n un sous ensemble d'environnements tel que $\lim_{n \rightarrow +\infty} \mathbb{P}(\mathcal{E}_n) = 1$ et*

$$\lim_{n \rightarrow +\infty} \inf_{\omega \in \mathcal{E}_n} \left\{ \mathbb{P}^\omega \left[\left| \frac{\mathcal{L}(n, m_{\log n})}{n} - \frac{1}{\mathbb{E}_{m_{\log n}}^\omega [\mathcal{L}(T_{m_{\log n}}, W_n)]} \right| \leq \frac{(\log_2 n)^{-1}}{\mathbb{E}_{m_{\log n}}^\omega [\mathcal{L}(T_{m_{\log n}}, W_n)]} \right] \right\} = 1.$$

La variable aléatoire $\mathbb{E}_{m_{\log n}}^\omega [\mathcal{L}(W_n, T_{m_{\log n}})]$ (le temps moyen quenched du temps dépensé par la marche dans la vallée W_n en un temps de retour à $m_{\log n}$) joue le rôle du temps moyen de retour en $m_{\log n}$ avant l'instant n . On retrouve ainsi le comportement du temps local d'une marche qui serait conditionnée à rester dans W_n , ici le conditionnement n'étant vrai que localement dans le temps.

Idées des preuves

En considérant la remarque ci-dessus, les preuves données dans [2] pourraient être simplifiées. Il me semble cependant intéressant de reprendre l'expression de $\mathbb{E}_{m_{\log n}}^\omega [\mathcal{L}(W_n, T_{m_{\log n}})]$ qui est la moyenne des temps d'excursion à droite et à gauche de $m_{\log n}$ dans la vallée W_n , on a

$$\mathbb{E}_{m_{\log n}}^\omega [\mathcal{L}(T_{m_{\log n}}, W_n)] \asymp \sum_{x \in W_n} e^{-(S_x - S_{m_{\log n}})}, \quad (1)$$

où \asymp signifie "encadré par, à une constante multiplicative près". On voit donc apparaître une somme d'exponentiel d'une différence de potentiel qui se trouve être toujours positive puisque $m_{\log n}$ est le $\log n$ -minima de la vallée W_n . Il est ainsi montré que la moyenne de $\mathbb{E}_{m_{\log n}}^\omega [\mathcal{L}(T_{m_{\log n}}, W_n)]$ est bornée. \square

Avec le recul on voit dans le résultat précédent et l'expression (1) les prémices de ce que sera la loi limite du supremum des temps locaux $\mathcal{L}^*(n) := \sup_{x \in \mathbb{Z}} \mathcal{L}(n, x)$ obtenu par N. Gantert, Y. Peres et Z. Shi, il s'écrit de la manière suivante :

Théoreme 1.4 ([GPS10]).

$$\frac{\mathcal{L}^*(n)}{n} \xrightarrow{\mathcal{L}} \sup_{x \in \mathbb{Z}} \pi(x) \quad (2)$$

où

$$\pi(x) := \frac{\exp(-S_x^\uparrow) + \exp(-S_{x-1}^\uparrow)}{2 \sum_{y \in \mathbb{Z}} \exp(-S_y^\uparrow)}, \quad x \in \mathbb{Z},$$

avec \mathbf{S}^\uparrow une suite de variables aléatoires distribuées comme \mathbf{S} conditionnée à rester strictement positive pour $x > 0$ et positive pour $x < 0$. $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi pour $n \rightarrow +\infty$.

Notons que le fait que le temps local pour la marche en milieu aléatoire récurrente peut être de l'ordre de n avait été remarqué initialement par Révész [R88].

Il est naturel de se poser la question d'une convergence plus forte pour la concentration de la marche, pour cela on peut définir la taille de concentration par la variable aléatoire suivante soit $0 \leq \beta \leq 1$

$$Y_{n,\beta} := \inf \left\{ k > 0, \sup_{x \in \mathbb{Z}} \sum_{j=x-k}^{x+k} \mathcal{L}(n, j) \geq \beta n \right\}.$$

Ainsi, dans [3], il est montré que \mathbb{P} -presque sûrement $\liminf Y_{n,\beta} \leq \text{const}(1 - \beta)^{-2}$. La preuve est basée sur une loi du zero-un pour les temps locaux obtenue dans Gantert-Shi [GS02], ainsi qu'une étude fine de la somme des excursions en dehors de $m_{\log n}$ de la marche dans un voisinage restreint de ce même point.

En fait à partir des résultats de [GPS10] où la limite supérieure du supremum des temps locaux est obtenue

Théoreme 1.5 ([GPS10]). $\mathbb{P}.p.s.$

$$\limsup_{n \rightarrow +\infty} \frac{\mathcal{L}^*(n)}{n} = c_1, \quad (3)$$

où c_1 est donnée explicitement comme fonction du support de la distribution de ω_0 ,

on peut améliorer le résultat de [3] sous la forme suivante

Théoreme 1.6 (A. [3],[5]). *Pour tout $0 \leq \beta \leq 1$,*

$$\mathbb{P} \text{ a.s. } \liminf_n Y_{n,\beta} = \begin{cases} 0 & \text{si } 0 \leq \beta \leq c_1, \\ f(\beta) & \text{si } c_1 < \beta < 1, \\ +\infty & \text{si } \beta = 1, \end{cases}$$

où f est connue explicitement.

Je ne donne pas l'expression de f ici, mais on peut retenir son comportement pour des β proche de 1 :

$$\mathbb{P} \text{ a.s. } \lim_{\beta \rightarrow 1^-} \liminf_n \frac{Y_{n,\beta}}{|\log(1-\beta)|} = c_2,$$

où $c_2 > 0$ est une constante.

Que dire de la limite supérieure ? Le résultat de la limite inférieure du supremum des temps locaux obtenu par A. Dembo, N. Gantert, Y. Peres et Z. Shi [DGPS07] suggère la chose suivante :

$$\mathbb{P} \text{ a.s. } \limsup_n \frac{Y_{n,\beta}}{\log \log \log n} = c_3(\beta).$$

On remarque que l'aspect concentration fait abstraction de la chronologie des événements mais tient compte de toute la trajectoire jusqu'à l'instant n , sous cet angle on pourrait penser que la marche est presque figée. Le résultat de concentration a ainsi été utilisé dans [Zin08] pour l'étude de la marche de Sinai en scénario aléatoire. Cependant notons que le lieu des points favoris n'est pas nécessairement dans le voisinage de concentration comme cela a été montré dans [SZ07]. Finalement la concentration n'est pas une conséquence directe du résultat de localisation (même si le travail de Sinai sert de base à son étude) mais n'implique pas non plus le résultat de localisation.

1.3 Prolongements

Suite à ces premiers travaux je me suis notamment orienté vers des applications des résultats obtenus pour la MAMA récurrente. Une partie concerne des problèmes d'estimation dont je parlerai dans la Section 4. Je me suis également intéressé à un système de particules de type Sinai, et à quoi pourrait ressembler une marche de Sinai sur \mathbb{Z}^d :

- **Système de particules en milieu aléatoire**

Dans [4], un système de particules indépendantes sur un milieu aléatoire est étudié : on distribue un nombre aléatoire (variables de Poisson i.i.d. de paramètre λ) de particules en chaque point du réseau \mathbb{Z} , chaque particule effectuant une marche de Sinai. Si $\eta(x, n)$ désigne le nombre de particules en $x \in \mathbb{Z}$ à l'instant n et f une fonction C^1 à support compact, j'obtiens la convergence en probabilité de $\frac{1}{(\log n)^2} (\sum_{x \in \mathbb{Z}} \eta(x, n) f(x/(\log n)^2))$. Pour aller plus loin il faudrait étudier les

cas où les particules sont en interaction, en s'inspirant des nombreux travaux sur ce domaine, citons par exemple E. Presutti, A. de Masi [DP91] et C. Kipnis C. Landim [KL98].

• **Potentiel de Sinai radial**

L'étude des MAMA sur \mathbb{Z}^d est beaucoup plus subtile (je n'en parlerai pas et renvoie donc à la référence usuelle [Zei01]), le travail [6] n'avait pas pour but d'étudier ces marches en toute généralité, mais plutôt de projeter l'idée de la marche de Sinai sur \mathbb{Z}^d . En faisant cela le plus simplement possible on a envie de définir un potentiel qui ferait en sorte que la marche se localise successivement sur des surfaces de \mathbb{Z}^d .

J'étudie donc une marche au plus proche voisin $(X_n, n \in \mathbb{N})$ réversible à valeurs dans \mathbb{Z}^d ($d > 1$), dont les probabilités de transition sont données par $p(x, y) = \exp(V(x) - V(y)) / [\sum_{z \sim x} \exp(V(x) - V(z))]$ si $y \sim x$, 0 sinon ($y \sim x$ signifiant que y est un plus proche voisin de x). A chacun des points de $x \in \mathbb{Z}^d$, V est donné par un potentiel de Sinai légèrement perturbé : $V(x) := S_{\|x\| + \delta_x}$. La norme $\|\cdot\|$ choisie ici est la norme infinie, et $(\delta_x, x \in \mathbb{Z}^d)$ sont des variables aléatoires à valeurs entières i.i.d bornées. Un phénomène de pseudo-localisation apparait : si $C_{m_{\log n}}$ est l'ensemble des points à une distance $m_{\log n}$ de l'origine, alors la probabilité que X_n soit dans le voisinage de $C_{m_{\log n}}$ converge vers un. On obtient également le comportement asymptotique en probabilité et en loi des temps locaux dans le voisinage de $C_{m_{\log n}}$ dont on peut déduire la limite en loi de $\max_k \mathcal{L}(C_k, n)/n$. Ce travail n'a finalement pas été publié, cependant l'idée d'étudier la façon dont la marche visite un ensemble de sites a été reprise pour des marches sur des arbres de Galton-Watson Section 3.

2 Temps et espace continu

Le processus étudié $(X(t), t \in \mathbb{R}_+; X(0) = 0)$ à valeurs dans \mathbb{R} est défini formellement par l'équation différentielle stochastique $dX(t) = d\beta(t) - 1/2V'(X(t))dt$, où le potentiel V est un processus càdlàg, et β un mouvement Brownien standard indépendant de V . Intuitivement on retrouve l'idée du cas discret de la Partie 1 : le processus X au voisinage de l'instant t et moyennant la fluctuation Brownienne infinitésimale $d\beta_t$ aura tendance à s'accroître si V est localement décroissant en $X(t)$. X a donc tendance à être attiré par les minima locaux de V . Cependant on ne peut définir rigoureusement le processus $(X(t), t)$ qu'à partir du générateur infinitésimal $\frac{1}{2}e^{V(x)}\frac{d}{dx}(e^{-V(x)}\frac{d}{dx})$ pour tout potentiel V fixé. Ainsi à potentiel fixé X s'écrit comme un mouvement Brownien changé de temps et d'espace, et c'est cette écriture qui est naturellement utilisée pour son étude. Comme auparavant \mathbb{P} désignera la mesure annealed.

Les premiers travaux apparaissent plus tardivement que pour la MAMA discrète, en particulier lorsque V est un mouvement Brownien, S. Schumacher [Sch85] puis T. Brox [Bro86] donnent les premiers arguments, suivis d'autres travaux comme ceux de H. Tanaka [Tan94], K. Kawazu et H. Tanaka [KT97], H. Tanaka [Tan97], puis Y. Hu, Z. Shi et M. Yor [HSY00].

Prenons donc $V = W_\kappa$ où pour tout $x \in \mathbb{R}$, $W_\kappa(x) := (W^{(1)}(-x) - \kappa x/2)\mathbf{1}_{x < 0} + (W^{(2)}(x) - \kappa x/2)\mathbf{1}_{x \geq 0}$, où $W^{(1)}$ et $W^{(2)}$ sont deux mouvements Brownien standards indépendants.

Comme je l'ai dit dans l'introduction, c'est l'article de Z. Shi [Shi98] qui a éveillé ma curiosité pour ces processus, voici son résultat, désignons par \mathcal{L}_X le temps local de X et $\mathcal{L}_X^*(t) := \sup_{x \in \mathbb{R}} \mathcal{L}_X(x, t)$ son supremum avant l'instant t .

Théoreme 2.1 ([Shi98]). *Supposons $\kappa = 0$, alors \mathbb{P} .p.s.*

$$\limsup_{t \rightarrow \infty} \frac{\mathcal{L}_X^*(t)}{t \log_3 t} \geq \frac{1}{32}. \quad (4)$$

Ce résultat est à comparer avec le résultat du cas discret (3), on voit donc que la curiosité vient de la normalisation. En effet alors que conceptuellement le cas discret (MAMA) et le cas continu (DMA) sont semblables, des différences inextricables liées à la nature des objets apparaissent pour certaines de leurs caractéristiques.

2.1 Le temps local du cas récurrent

On se place ici dans le cas $V = W_0$, on appellera X processus de Brox en référence à T. Brox [Bro86] qui montre que le comportement asymptotique (en loi) de X à l'instant t est le même que pour la marche de Sinai. On peut noter qu'il en va de même pour les limites presque sûres comme le montrent Z. Shi et Y. Hu dans [HS98].

La curiosité dont parle Z. Shi n'apparaît donc pas (de cette façon là en tout cas) dans le comportement du processus X à un instant donné mais dans le temps local.

L'étude va donc porter sur le temps local $(\mathcal{L}_X(t, x), t \in \mathbb{R}_+, x \in \mathbb{R})$ de X défini par :

$$\int_0^t f(X_s)ds = \int_{\mathbb{R}} f(x)\mathcal{L}_X(t, x)dx,$$

pour toute fonction mesurable f . On voit immédiatement que le temps local pour X est une densité qui n'a, a priori, aucune raison d'être plus petite que t contrairement au temps local de la marche discrète où $\mathcal{L}(n, x) \leq n$, $x \in \mathbb{Z}$. Pour améliorer le résultat (4), une option naturelle était de reprendre les travaux de [GPS10] pour le cas discret et les transposer au cas continu. Malheureusement (ou heureusement d'ailleurs!), on ne va pas pouvoir passer aussi facilement

de la loi limite du supremum des temps locaux à sa la limite supérieure.

Avec R. Diel que j'ai co-encadré avec R. Abraham, nous avons donc étendu une partie des résultats de Gantert-Peres-Shi [GPS10] pour X . L'étude va donc être basée sur l'estimation du temps local au voisinage du point de localisation. Pour cela nous avons aussi besoin de la notion de h -extrema et h -vallée dans le cas continu : pour tout $h > 0$, $x \in \mathbb{R}$ est un h -minimum pour un processus V si il existe $u < x < v$ tel que $V(y) \geq V(x)$ pour tout $y \in [u, v]$, $V(u) \geq V(x) + h$ et $V(v) \geq V(x) + h$. Aussi x est un h -maximum pour V si x est un h -minimum pour $-V$, x est un h -extrema pour V si c'est un h -maximum ou un h -minimum pour V . Une h -vallée ici suit la même notion que dans le cas discret. Notons $(W_\kappa^\uparrow(x), x \geq 0)$ et $(\widetilde{W}_\kappa^\uparrow(x), x \geq 0)$ deux copies indépendantes du processus $(W_\kappa(x), x \geq 0)$ conditionné à rester positif. Ici $\kappa = 0$, $(W_\kappa^\uparrow(x), x \geq 0)$ et $(\widetilde{W}_\kappa^\uparrow(x), x \geq 0)$ sont tous deux égaux en loi à un processus de Bessel de dimension 3 partant de 0. Définissons alors la fonctionnelle :

$$\mathcal{R}_\kappa := \int_0^{+\infty} e^{-W_\kappa^\uparrow(x)} dx + \int_0^{+\infty} e^{-\widetilde{W}_\kappa^\uparrow(x)} dx,$$

et $0 < \mathcal{R}_\kappa < +\infty$ presque sûrement. On a

Théoreme 2.2 (A.-Diel [11]). *Supposons $\kappa = 0$, alors*

$$\frac{\mathcal{L}_X^*(t)}{t} \xrightarrow{\mathcal{L}} \frac{1}{\mathcal{R}_0}.$$

On voit que l'idée du Théorème 1.4 demeure intacte ici pour le processus de Brox. En re-normalisant par le temps imparti pour le processus, on obtient la convergence en loi de $\mathcal{L}_X^*(t)$ vers l'inverse de la somme de deux fonctionnelles exponentielles indépendantes du potentiel conditionné à rester positif.

Idées de preuve. Notons que dans [11] la propriété de scaling des mouvements Browniens en jeux est largement utilisée comme dans [Bro86] mais n'est pas nécessaire, nous donnons ici les idées sans cela.

La preuve de ce résultat passe d'une part par l'étude, aujourd'hui classique, de la vallée-prison, c'est à dire essentiellement l'étude des fluctuations du processus recentré $(\Delta W_{m_{\log t}}(x) := W_0(x) - W_0(m_{\log t}), x \in [M_{\log t}^-, M_{\log t}^+])$, où $m_{\log t}$ est le $\log t$ -minima de W_0 le plus proche de 0, $M_{\log t}^-$ (resp. $M_{\log t}^+$) la coordonnée la plus proche (à droite avec le $+$ et à gauche avec le $-$) de $m_{\log t}$ tel que $W_0(M_{\log t}^\pm) - W_0(m_{\log t}) = \log t$. Et d'autre part, par l'étude du temps local recentré en $m_{\log t}$: après avoir montré que le processus atteint $m_{\log t}$ en un temps négligeable devant t , on étudie l'inverse du temps local en $m_{\log t}$, (ce qui implique un résultat équivalent à la Proposition 1.3) c'est ici qu'apparaît une fonctionnelle exponentielle qui s'écrit $R_0(t) = \int_{M_{\log t}^-}^{M_{\log t}^+} e^{-\Delta W_{m_{\log t}}(x)} dx$. On estime ensuite le temps local dans un petit voisinage

$I_t \subset [M_{\log t}^-, M_{\log t}^+]$ de $m_{\log t}$ (tel qu'en dehors de ce voisinage le temps local est négligeable), jusqu'à l'inverse du temps local en $m_{\log t}$. Ces deux résultats induisent le comportement asymptotique du temps local normalisé par t pour tous les points de I_t , ce processus étant ainsi proche de $(e^{-\Delta W_{m_{\log t}}(x)}/R_0(t), x \in I_t)$ en probabilité. Remarquons qu'il est naturel ici de retrouver l'expression de la mesure invariante d'un processus analogue au processus X_t réfléchi en $M_{\log t}^-$ et $M_{\log t}^+$ comme nous en avons discuté pour la localisation de la marche de Sinai. En utilisant à la fois que $(\Delta W_{m_{\log t}}(-x), -m_{\log t} \leq x \leq -M_{\log t}^-)$ et $(\Delta W_{m_{\log t}}(x), m_{\log t} \leq x \leq M_{\log t}^+)$ sont indépendants et égaux en loi à un processus de Bessel de dimension 3 partant de 0 arrêtés en $\log t$, et un résultat de Tanaka [Tan94], nous obtenons la convergence en loi (pour la topologie de Skorohod usuelle J_1) du processus recentré $(\mathcal{L}_X(t, x + m_{\log t})/t)_{x \in \mathbb{R}}$ vers le processus

$\left((e^{-W_0^\uparrow(x)} \mathbf{1}_{x \leq 0} + e^{-\widetilde{W}_0^\uparrow(x)} \mathbf{1}_{x > 0}) / \mathcal{R}_0 \right)_{x \in \mathbb{R}}$. Pour déterminer la loi limite du supremum des temps locaux on doit travailler un peu plus que dans le cas discret en repérant les points favoris, on utilise donc le résultat de Cheliotis [Che08] qui montre la proximité de $m_{\log t}$ et du plus petit (en valeur absolu) point favori de X . \square

Ce résultat se généralise au cas où V est un processus de Lévy α -stable $\alpha \in [1, 2]$ qui ne soit pas une dérive pure, voir pour cela R. Diel et G. Voisin [Die11b].

Pour obtenir les convergences presque sûres, une étude plus fine, à la fois sur le milieu et le comportement du processus, était nécessaire. R. Diel dans son travail de thèse a amélioré le résultat de curiosité de Z. Shi de la façon suivante

Théoreme 2.3 (Diel [Die11a]). *Supposons $\kappa = 0$, ils existent des constantes $0 < c_1^\pm < c_2^\pm < +\infty$ telles que \mathbb{P} .p.s.*

$$c_1^+ \leq \limsup_{t \rightarrow +\infty} \frac{\mathcal{L}_X^*(t)}{t \log_3(t)} \leq c_2^+, \quad c_1^- \leq \liminf_{t \rightarrow +\infty} \frac{\log_3(t) \mathcal{L}_X^*(t)}{t} \leq c_2^-.$$

L'idée de Diel pour la preuve est d'abandonner la localisation dans le voisinage d'un unique h -minima. Il étudie donc 4 vallées (deux à gauche et deux à droite de l'origine) dont les profondeurs sont légèrement supérieures et inférieures à $\log t$. Il montre alors que presque sûrement le processus va passer la quasi totalité de son temps dans les voisinages de ces minima dont les tailles sont inférieures à $(\log \log t)^{4+\epsilon}$. Cette approche est un mélange entre les résultats de [SZ07] qui montrent que le point favori de la marche peut se trouver éloigné de la région où la marche dépense la quasi totalité de son temps et de [3] qui montre que la marche peut dépenser la quasi totalité de son temps dans une région très restreinte de \mathbb{Z} . Il remarque aussi que la variable clé qui donne cette subtilité dans les normalisations est celle que nous avons notée \mathcal{R}_0 . C'est la différence de nature entre cette variable et la somme $2 \sum_{y \in \mathbb{Z}} e^{-S_y^\uparrow}$ qui apparaît dans le Théoreme 1.4 qui induit la curiosité de Shi : \mathbf{S}^\uparrow et W_κ^\uparrow (ainsi que $\widetilde{W}_\kappa^\uparrow$) sont tous deux obtenus à partir du potentiel qui est conditionné à rester positif, tous deux peuvent être relativement plats de la même façon, ce qui donne des limites inférieures identiques pour le cas discret et continu (voir [DGPS07] pour le cas discret) mais W_κ^\uparrow autorise des vallées extrêmement étroites faisant apparaître le $\log \log \log t$ supplémentaire dans la limite supérieure du cas continu.

▷ Poursuivre sur le temps local dans cette direction impliquait aussi d'étudier le cas transient. Pour cela il est bien de parler un peu du cas (transient) discret, reprenons donc les notations de la première section. On suppose que (X_n, n) est une marche telle que $\mathbb{E}(\log[(1 - \omega_0)/\omega_0]) < 0$ qui implique la transience à droite et qu'il existe $\kappa > 0$ tel que $\mathbb{E}([(1 - \omega_0)/\omega_0]^\kappa) = 1$. Rappelons que les comportements asymptotiques pour ces marches ont été initialement déterminés par Kesten-Kozlov-Spitzer [KKS75], ainsi X_n présente une multitude de comportements (pas uniquement polynomiale en n , ce qui en fait sa richesse) en fonction de κ . Le temps local et plus particulièrement le supremum des temps locaux a été étudié par Gantert-Shi [GS02]. Nous ne parlerons pas ici du cas $\kappa \geq 1$, bien que des résultats soient également obtenus pour ces cas dans ce même article. Voici le résultat qui est notre point de départ

Théoreme 2.4 ([GS02]). *Si $0 < \kappa < 1$, alors il existe $c > 0$ tel que \mathbb{P} -p.s.*

$$\limsup_{n \rightarrow +\infty} \frac{\mathcal{L}^*(n)}{n} = c.$$

Gantert-Shi font remarquer que pour ce cas transient la marche peut passer énormément de temps en un unique point, comme le fait la marche de Sinai. La méthode pour obtenir le

résultat est basée (voir [KKS75]) sur un lien fort entre le processus des nombres de pas à gauche arrêté et un processus de branchement en milieu aléatoire avec immigrant. Je reparlerai un peu de ceci dans la Section 4, puisque cela nous sera utile à ce moment. Dans la continuité de ce résultat et des résultats dont nous avons parlé précédemment, je me suis intéressé au pendant continu de (X_n, n) . \triangleright

2.2 Localisation du cas transient lent

Je continue en ce début de section à parler du cas discret, et en particulier des travaux de Enriquez-Sabot-Zindy qui ont été une source d'inspiration importante pour nous.

Commençons par une remarque générale sur la façon dont les résultats sur le comportement asymptotique des MAMA ont été obtenus de part et d'autre. Pour la marche de Sinai, c'est une description quenched de l'environnement qui a permis la caractérisation du point de localisation qui a lui-même conduit à la détermination de la loi limite. Il n'en est pas de même pour la marche transiente dans les premiers travaux [KKS75]. Il faut ainsi attendre ceux de N. Enriquez, C. Sabot et O. Zindy pour voir apparaître la démarche de Sinai adaptée aux cas transients. Je me contenterai ici d'étudier et discuter uniquement le cas $0 < \kappa < 1$. Dans une série de 3 travaux [ESZ09c],[ESZ09b],[ESZ09a] Enriquez-Sabot-Zindy montrent (entre autre) l'existence d'une localisation de la marche à l'instant n et obtiennent une caractérisation complète du point de localisation. Prenons $h_n = \log n - \log \log n$, et notons $m_{h_n}^{(i)}$ le i -ième h_n -minima positif, leur résultat de localisation s'écrit

Théoreme 2.5 ([ESZ09a]). *Supposons $0 < \kappa < 1$, pour tout $\epsilon > 0$*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(|X_n - m_{h_n}^{(N_n)}| \leq \epsilon \log n \right) = 1,$$

où $N_n := \sup\{i \geq 1, \sup_{1 \leq k \leq n} X_k \geq m_{h_n}^{(i)}\}$.

La localisation dans ces cas transients lents est donc, en partie, semblable au cas Sinai : on peut trouver un h -minima pour lequel X_n se retrouve dans son voisinage à l'instant n . Cependant on ne peut pas trouver pour h une valeur qui serait à l'origine d'une unique vallée, contenant l'origine, arrêtant définitivement la marche avant l'instant n . Ainsi la marche utilise son temps à s'extraire d'une succession de h_n -minima semblables et termine sa course dans une ultime vallée dont les caractéristiques sont semblables aux précédentes par construction (c'est encore un h_n -minima). Nous verrons dans la section 2.3 que cette dernière vallée est cependant stochastiquement un peu différente.

Même si l'objectif premier était d'étudier la loi limite du supremum des temps locaux renormalisé, ils nous a semblé bien avec A. Devulder de commencer par étendre ce résultat de localisation au cas continu.

On se donne une hauteur de vallée fixée $h_t = \log t - \phi(t)$, où ϕ est une fonction croissante positive typiquement $\phi(t) = o(\log t)$ mais $\lim_{t \rightarrow +\infty} \phi(t)/(\log \log t) = +\infty$. Comme pour le cas discret on note $m_{h_t}^{(i)}$ le i -ième h_t -minima positif ($i \geq 1$), par le travail de [Fag09] nous savons que presque sûrement les h_t -extrema de W_κ forment une sous suite indexée par \mathbb{Z} non bornée, et que les h_t -minima et h_t -maxima alternent. Le résultat s'écrit alors de façon très similaire à celui de [ESZ09a]. Notons $N_t := \sup\{i \geq 1, \sup_{s \leq t} X(s) \geq m_{h_t}^{(i)}\}$, nous montrons qu'à l'instant t , X est localisé dans le voisinage du dernier h_t -minima rencontré :

Théoreme 2.6 (A.-Dev [10]). *Supposons $0 < \kappa < 1$, il existe une constante $\mathcal{C}_1 > 0$, telle que*

$$\lim_{t \rightarrow +\infty} \mathbb{P} \left(|X(t) - m_{h_t}^{(N_t)}| \leq \mathcal{C}_1 \phi(t) \right) = 1.$$

La précision que l'on obtient ici est semblable à celle obtenue par [BF08] (Théorème 3) pour la marche de Sinai, et dépend de h_t . Comme nous le voyons le résultat dépend fortement de N_t qui dépend lui même de X , nous en savons un peu plus sur ce nombre de h_t -vallées visitées jusqu'à l'instant t par le résultat suivant

Proposition 2.7 (A.-Dev [10]). $N_t e^{-\kappa\phi(t)}$ converge en loi sous la mesure annealed \mathbb{P} vers \mathcal{N} qui est déterminée par sa transformée de Laplace :

$$\forall u > 0, \quad \mathbb{E}(e^{-u\mathcal{N}}) = E_{\kappa,1}(-u/C_\kappa),$$

où $E_{\kappa,1}()$ est la fonction de Mittag-Leffler de paramètres $(\kappa, 1)$ et $C_\kappa > 0$ est connue explicitement.

Rappelons ici que [KT97] montre que $X(t)/t^\kappa$ converge en loi, leur résultat peut être obtenu à partir du théorème et de la proposition ci-dessus moyennant un petit travail supplémentaire que nous évoquerons dans la partie prolongement.

Idées des preuves

Plusieurs idées apparaissent dans la preuve, une première partie où on construit une suite de h_t -minima spécifiques qui associée au fait que le processus revient peu sur ses pas induit de l'indépendance dans la trajectoire de X . Une seconde partie où on étudie les temps d'extraction des $n_t := \lfloor e^{\kappa\phi(t)(1+\delta)} \rfloor$ ($\delta > 0$, choisi petit) premiers h_t -minima. Ceci fait apparaître une structure de renouvellement pour le temps d'atteinte de la dernière h_t -vallée avant l'instant t . Enfin une troisième partie où l'on se sert du renouvellement et d'un argument similaire à celui utilisé pour le processus de Brox pour montrer la localisation.

• La suite de h_t -minima pour le Brownien drifté (W_κ) a été étudiée par A. Faggionato [Fag09]. Ce travail permet, en particulier, d'avoir une décomposition de type Williams dans le voisinage des h_t -minima. Cette décomposition se reflète dans les fonctionnelles définies un peu plus loin dans la discussion.

On construit alors une suite de h_t -minima notés $(\tilde{m}_{h_t}^{(i)}, i \leq n_t)$ qui ont la propriété suivante : la suite $\{(\Delta W_\kappa^{(i)}(x) := W_\kappa(x) - W_\kappa(\tilde{m}_{h_t}^{(i)}), \tilde{L}_i^- \leq x \leq \tilde{L}_i^+, i \leq n_t)\}$ [où \tilde{L}_i^- et \tilde{L}_i^+ sont représentés sur la Figure 1 et correspondent à des coordonnées qui dépassent assez largement le puits de potentiel créé autour de $\tilde{m}_{h_t}^{(i)}$] est composée de processus i.i.d. On a aussi besoin de la coordonnée $\tilde{L}_i < \tilde{L}_i^+$ située après le i ème h_t -maximum $\tilde{M}_{h_t}^{(i)}$ et tel que $\Delta W_\kappa^{(i)}(\tilde{L}_i) = h_t/2$. On montre alors qu'avec une grande probabilité les suites $(\tilde{m}_{h_t}^{(i)}, i \leq n_t)$ et $(m_{h_t}^{(i)}, i \leq n_t)$ coïncident. Sur un événement de probabilité proche de 1, on peut ainsi profiter à la fois de l'indépendance et de la décomposition de Williams.

On montre aussi qu'une fois en un minima $\tilde{m}_{h_t}^{(i)}$ donné le processus touchera le prochain h_t -minima avant de revenir à la coordonnée \tilde{L}_i^- , et ceci pour les n_t premiers h_t -minima. Si on regarde maintenant le temps qu'il faut au processus pour passer d'un minima local au suivant $(H(\tilde{m}_{h_t}^{(i+1)}) - H(\tilde{m}_{h_t}^{(i)}))^1$, on montre que ce temps est proche de $H_i := H(\tilde{L}_i) - H(\tilde{m}_{h_t}^{(i)})$ avec X conditionné à rester en dessous de \tilde{L}_i^- . Ceci impliquant que

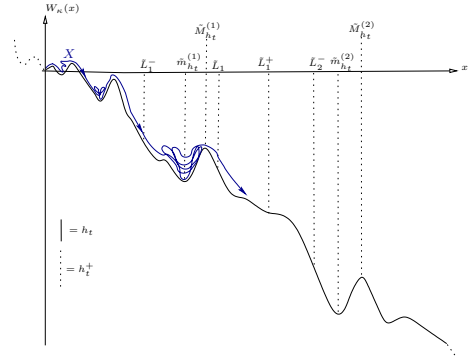


FIGURE 1 – Potentiel - W_κ

1. $H(x) := \inf\{s > 0, X(s) = x\}$, $x \in \mathbb{R}$

cette partie de la trajectoire dépend uniquement de $(\Delta W_\kappa^{(i)}(x), \tilde{L}_i^- \leq x \leq \tilde{L}_i)$. Si on ajoute à cela le fait que les temps entre les h_t -vallées² $(H(\tilde{m}_{h_t}^{(i+1)}) - H(\tilde{L}_i), i)$ sont négligeables, le temps t'atteinte d'un h_t -minima quelconque $j \leq n_t$ s'écrit en probabilité comme une somme de variables aléatoires $(H_i, i \leq j-1)$ indépendantes et de même loi.

• L'étape suivante est donc d'étudier la v.a. H_1 . C'est cette partie où le fait d'étudier le processus $X(t)$ plutôt que la marche de Kesten-Kozlov-Spitzer, rend les choses un peu plus simple sans pour autant supprimer la technicité. On montre dans un premier temps que H_1 peut s'écrire (en probabilité) sous la forme d'un produit de trois variables aléatoires indépendantes, définissons

$$F^\pm(h_t) := \int_0^{\tau_{W_\kappa^\uparrow}(h_t)} e^{\pm W_\kappa^\uparrow(x)} ds, \quad G^+(h_t, h_t/2) := e^{h_t} \int_0^{\tau_{\tilde{W}_\kappa}(-h_t/2)} e^{\tilde{W}_\kappa(x)} ds.$$

où pour un processus continu U , et $a \in \mathbb{R}$, $\tau_U(a) := \inf\{s > 0, U = a\}$, \tilde{W}_κ est une copie de W_κ et W_κ^\uparrow le potentiel conditionné à rester positif défini au dessus du Théorème 2.2. On montre alors qu'en probabilité $|H_1 - \mathcal{H}_1| \leq \epsilon_t \mathcal{H}_1$ (avec $\epsilon_t \rightarrow 0$), où

$$\mathcal{H}_1 := R_1 S_1 \mathbf{e}_1,$$

avec $R_1 \stackrel{\mathcal{L}}{=} F_1^-(h_t) + \tilde{F}_1^-(h_t)$, $S_1 \stackrel{\mathcal{L}}{=} F_1^+(h_t) + G_1^+(h_t, h_t/2)$ et $\mathbf{e}_1 \sim \mathcal{E}(1/2)$ des variables aléatoires indépendantes. De plus $F_1^-(h_t)$ et $\tilde{F}_1^-(h_t)$ sont des copies indépendantes de $F^-(h_t)$, $F_1^+(h_t)$ une copie indépendante de $F^+(h_t)$ et $G_1^+(h_t, h_t/2)$ une copie de $G^+(h_t, h_t/2)$ indépendante de $F_1^+(h_t)$. Notre approche est un peu différente que dans le cas discret [ESZ09c] et découle en partie de la décomposition de type Williams obtenue par [Fag09]. Une seconde différence vient du fait que nous nous intéressons uniquement à des fonctionnelles tronquées (à certains temps d'arrêt). Rappelons en effet que le travail [ESZ09c] n'a pas pour unique objectif l'estimation du temps de sortie H_1 , mais en est plutôt une conséquence (voir Colloraire 7.1 et Remarque 7.1 de leur appendice).

On étudie ensuite la queue de distribution de \mathcal{H}_1/t , la re-normalisation par t étant intrinsèque au problème que l'on s'est posé : étudier le processus jusqu'à l'instant t dont on sait qu'il est découpé en somme de variables i.i.d. de même loi que \mathcal{H}_1 (les contributions en temps en dehors de ces temps de sortie étant négligeable). On calcule pour cela le second ordre de la transformée de Laplace de \mathcal{H}_1 qui donne le résultat suivant, pour t grand

$$e^{\kappa\phi(t)} \left(1 - \mathbb{E} \left(e^{-\frac{\lambda}{t} \mathcal{H}_1} \right) \right) = C_\kappa \lambda^\kappa + o(1).$$

Ici aussi le cas continu simplifie un peu le problème par rapport au cas discret (voir Sections 5, 6 et 7 dans [ESZ09b]), car les transformées de Laplace des variables $F^\pm(h_t)$ et $G(h_t, h_t/2)$ sont connues. Cependant étant donné que l'on étudie les produits de ces variables il faut composer les transformées de Laplace, et le fait qu'elles s'écrivent comme fraction de fonctions de Bessel modifiées, implique un peu de technicité pour obtenir le résultat. On obtient donc ensuite la queue de distribution à droite de \mathcal{H}_1/t : par un théorème Taubérien, $e^{-\kappa\phi(t)} \mathbb{P}(\mathcal{H}_1/t \geq x) \sim x^{-\kappa} C_\kappa / \Gamma(1 - \kappa)$.

Les deux premiers points de la preuve permettent donc de transposer certains problèmes relatifs aux temps d'atteinte des h_t -minima à un problème de somme de variables aléatoires i.i.d à queue lourde. Par exemple l'étude de $\mathcal{N}_t := \max\{k \geq 1, \sum_{i=1}^k \mathcal{H}_i \leq t\}$ va permettre de déterminer la transformée de Laplace de N_t . La "traduction" d'un problème vers l'autre implique cependant de prendre certaines précautions car certains passages à la limite (en t) sont relativement délicats.

2. Nous entendons ici par h_t -vallée l'intervalle $(\tilde{L}_i^-, \tilde{L}_i)$ cela diffère légèrement de la définition initiale, en effet $\tilde{L}^{(i)}$ n'est pas une coordonnée de maximum sur l'intervalle considéré.

• Les résultats autour de la structure de renouvellement sont donc intéressants en soi indépendamment de la localisation. Cependant la bonne connaissance de la loi asymptotique de le $H(m_{h_t}^{(N_t+1)})/t$ est déterminante pour montrer la localisation. L'idée reprend aussi celle de Brox [Bro86] moyennant le fait que la localisation n'a pas lieu dans le h_t -minima le plus proche de 0. Le fait que plusieurs h_t -vallées (environ $e^{\kappa\phi(t)}$) soient impliquées pourrait créer une certaine instabilité dans le processus, si par exemple d'un instant t à un instant $t(1+\epsilon)$ avec ϵ petit le processus changeait de vallée il serait difficile de localiser le processus à l'instant t . Or par le résultat de renouvellement de type Dynkin ([Dyn55]-[Fel71] chapter XIV) que nous obtenons pour $H(m_{h_t}^{(N_t+1)})$

$$\lim_{t \rightarrow +\infty} \mathbb{P}\left(H(m_{h_t}^{(N_t+1)})/t \geq 1+s\right) = \frac{\sin(\pi\kappa)}{\pi} \int_s^{+\infty} (1+x)^{-1} x^{-\kappa} dx,$$

on a $\lim_{t \rightarrow +\infty} \mathbb{P}(N_t \neq N_{t(1+\epsilon)}) \leq \epsilon^{1-\kappa}$. On a donc une certaine stabilité pour N_t , la suite est alors routinière : la probabilité de trouver X à l'instant t dans le voisinage du minimum local se comporte grosso-modo comme la mesure invariante centrée en $m_{h_t}^{(N_t)}$ (du processus conditionné à rester dans la vallée que l'on aura couplé avec X pour ses derniers instants avant t) de ce voisinage. L'obtention du $\phi(t)$ dans le théorème nécessite de bien conserver les vitesses de convergence des probabilités tout au long des calculs. \square

Une conséquence de la localisation est le théorème de vieillissement, également obtenu dans le cas discret dans [ESZ09a] :

Proposition 2.8 (A.-Dev [10]). *Supposons $0 < \kappa < 1$, pour tout $\alpha > 1$*

$$\lim_{t \rightarrow +\infty} \mathbb{P}(|X(\alpha t) - X(t)| \leq \phi(t)) = \frac{\sin(\kappa\pi)}{\pi} \int_0^{1/\alpha} u^{\kappa-1} (1-u)^{-\kappa} du.$$

▷ Ce premier travail nous a permis de transposer une partie des résultats d'Enriquez-Sabot-Zindy au cas continu et ainsi d'apprécier les différences avec le cas discret, tout en cultivant ce qui serait nécessaire pour la problématique du temps local, et à l'extension à des milieux plus généraux. Le travail suivant a donc pour objectif de donner un premier résultat sur le temps local de X toujours dans le cas $V = W_\kappa$. Grégoire Véchambre que je co-encadre avec Romain Abraham, nous a rejoint sur ce problème. Nous allons voir que l'étude du supremum du temps local, telle que nous l'avons envisagée, va demander l'étude jointe du processus des temps de sortie et du temps local au niveau des h_t -minima. \triangleright

2.3 Le temps local du cas transient lent

La motivation première ici est l'étude du supremum des temps locaux pour $0 < \kappa < 1$, le découpage en vallées présenté précédemment sera de nouveau utilisé ici. Intuitivement on sait que le supremum du temps local avant l'instant t sera sans doute atteint dans le voisinage d'un ou plusieurs h_t -minima visités par X . On sait aussi que la structure de renouvellement a un rôle intrinsèque au comportement du processus. En particulier le nombre typique de vallées visitées est de l'ordre de $e^{\kappa\phi_t}$, et le temps d'atteinte de la k ème vallée est proche d'une somme de $k-1$ variables aléatoires indépendantes et identiquement distribuées. On sait aussi que les temps de sortie s'écrivent comme une fonctionnelle additive de temps locaux, et qu'ainsi les temps de sortie et les temps locaux pour un h_t -minima donné ont une partie commune. On va donc étudier le processus joint : somme des temps locaux et somme des temps de sorties en les $(\lfloor se^{\kappa\phi_t} \rfloor, s > 0)$ premiers h_t -minima. On va obtenir en limite un processus de Lévy qui contiendra toute l'information nécessaire à expliciter la loi limite de $\mathcal{L}^*(t)/t$.

Pour cela nous noterons $(D([0, +\infty), \mathbb{R}^d), J_1)$, $d > 0$ l'espace des fonctions càdlàg muni de la topologie de Skorohod J_1 et $\xrightarrow{\mathcal{L}_s}$ la convergence en loi des processus pour cette topologie. Soit $(\mathcal{Y}_1, \mathcal{Y}_2)$ un processus de Lévy caractérisé par sa mesure de Lévy ν donnée par

$$\forall x > 0, y > 0, \nu([x, +\infty) \times [y, +\infty)) = \frac{1}{y^\kappa} \mathbb{E} \left[(\mathcal{R}_\kappa)^\kappa \mathbb{1}_{\mathcal{R}_\kappa \leq \frac{y}{x}} \right] + \frac{1}{x^\kappa} \mathbb{P} \left(\mathcal{R}_\kappa > \frac{y}{x} \right),$$

où \mathcal{R}_κ est défini au dessus du Théorème 2.2. Notons aussi que \mathcal{R}_κ est la variable aléatoire limite, lorsque t tend vers l'infini, de la variable R_1 vue au paragraphe précédent.

Pour une fonction donnée f dans $D([0, +\infty), \mathbb{R})$, notons pour tout $s > 0, a > 0$:

$$f^\natural(s) := \sup_{0 \leq r \leq s} (f(r) - f(r^-)), \quad f^{-1}(a) := \inf\{x > 0, f(x) > a\},$$

$f^\natural(s)$ est le plus grand saut de f avant l'instant s , $f^{-1}(a)$ l'inverse f . Nous notons également $f^{-1}(a)^-$ le dernier instant où f est en dessous de a . Définissons maintenant le couple de variables aléatoires $(\mathcal{I}_1, \mathcal{I}_2)$

$$\mathcal{I}_1 := \mathcal{Y}_1^\natural(\mathcal{Y}_2^{-1}(1)^-), \mathcal{I}_2 := (1 - \mathcal{Y}_2(\mathcal{Y}_2^{-1}(1)^-)) \times \frac{\mathcal{Y}_1(\mathcal{Y}_2^{-1}(1)) - \mathcal{Y}_1(\mathcal{Y}_2^{-1}(1)^-)}{\mathcal{Y}_2(\mathcal{Y}_2^{-1}(1)) - \mathcal{Y}_2(\mathcal{Y}_2^{-1}(1)^-)}.$$

On a alors le résultat suivant

Théorème 2.9 (A.-Devulder-Vechambre [15]). *Pour $0 < \kappa < 1$,*

$$\frac{\mathcal{L}_X^*(t)}{t} \xrightarrow{\mathcal{L}} \mathcal{I} = \max(\mathcal{I}_1, \mathcal{I}_2).$$

L'interprétation du résultat est le suivant : pour tout $s > 0$, $\mathcal{Y}_1(s)$ doit être vu comme le processus limite qui correspond à la somme des $\lfloor se^{\kappa\phi(t)} \rfloor$ premiers temps locaux normalisés par t au niveau des $\lfloor se^{\kappa\phi(t)} \rfloor$ premiers h_t -minima. \mathcal{Y}_2 est l'équivalent (c'est à dire la somme) pour les temps de sorties des $\lfloor se^{\kappa\phi(t)} \rfloor$ premières h_t -vallées. Ainsi \mathcal{I}_1 est le plus grand saut du processus \mathcal{Y}_1 avant que \mathcal{Y}_2 ne dépasse 1, c'est à dire le plus grand temps local re-normalisé dans le fond des vallées dont le processus X est sorti avant l'instant t . \mathcal{I}_2 est un produit de deux termes. Le premier $(1 - \mathcal{Y}_2(\mathcal{Y}_2^{-1}(1)^-))$ est la différence entre 1 et la valeur de \mathcal{Y}_2 avant l'ultime saut qui lui fera dépasser la valeur 1, il correspond donc au temps (normalisé par t) consacré à la dernière vallée avant l'instant t par X . Pour le second terme (sous la forme d'une fraction), on peut remarquer que \mathcal{Y}_2 peut être construit à partir de \mathcal{Y}_1 en multipliant chacun des sauts de ce dernier par une copie (indépendante) de la variable \mathcal{R}_κ . Ainsi ce second terme est une copie de la variable $1/\mathcal{R}_\kappa$ prise à l'instant où \mathcal{Y}_2 fait le saut qui lui fera dépasser la valeur 1. Cette dernière variable joue un rôle similaire à la variable $1/\mathcal{R}_0$ du Théorème 2.2 puisque le processus X ici ne s'échappera pas de la dernière vallée. \mathcal{I}_2 est donc une proportion (du temps restant avant l'instant t) du temps local en l'ultime h_t -minima visité avant l'instant t .

Idées de preuve

Une première partie reprend les idées et outils nécessaires à la preuve de la localisation que l'on généralise, on étudie en effet le processus joint $(\sum \text{temps de sortie}, \sum \text{temps locaux en les } \tilde{m}_{h_t}^{(j)})$, sans tenir compte dans un premier temps que le processus est regardé jusqu'au temps t . Pour cela on commence par montrer qu'avec une grande probabilité $(H_j, \mathcal{L}(H(\tilde{L}_j), \tilde{m}_{h_t}^{(j)}); j \leq n_t)$ peut être approchée par une suite de variables aléatoires i.i.d, plus précisément :

Lemma 2.10. Soit $t > 0$, ils existent $c_1 > 0$ et une suite $(\{S_j(t), R_j(t), \mathbf{e}_j(t)\}, j \leq n_t)$ de variables aléatoires indépendantes et identiquement distribuées telle que S_j, R_j et \mathbf{e}_j sont également indépendantes avec $S_1 \stackrel{\mathcal{L}}{=} F_1^+(h_t) + G_1^+(h_t, h_t/2)$, $R_1 \stackrel{\mathcal{L}}{=} F_1^-(h_t/2) + \tilde{F}_1^-(h_t/2)$ et $\mathbf{e}_1 \stackrel{\mathcal{L}}{=} \mathcal{E}(1/2)$ tel que

$$\lim_{t \rightarrow +\infty} \mathbb{P} \left(\cap_{j=1}^{n_t} \{ |H_j - \mathcal{H}_j| \leq \epsilon_t \mathcal{H}_j, |\mathcal{L}(H(\tilde{L}_j), \tilde{m}_{h_t}^{(j)}) - \ell_j| \leq \epsilon_t \ell_j \} \right) = 1, \text{ avec}$$

$$\ell_j := S_j \mathbf{e}_j, \quad \mathcal{H}_j := R_j \ell_j, \quad \epsilon_t := e^{-c_1 h_t}.$$

De plus $e^{\kappa\phi(t)} \mathbb{P}(\ell_1/t \geq x, \mathcal{H}_1/t \geq y)$ converge quand t tend vers l'infini uniformément sur tout compact de $(0, +\infty) \times (0, +\infty)$ vers $\nu([x, +\infty) \times [y, +\infty))$.

On montre ensuite la convergence (lorsque t tend vers l'infini) du processus $(Y_1, Y_2)^t$ défini par

$$\forall s \geq 0, (Y_1, Y_2)_s^t := \frac{1}{t} \sum_{j=1}^{\lfloor s e^{\kappa\phi(t)} \rfloor} (\ell_j, \mathcal{H}_j),$$

vers le subordonateur défini précédemment $[(Y_1, Y_2)^t \xrightarrow{\mathcal{L}_S} (\mathcal{Y}_1, \mathcal{Y}_2)]$. L'approche que nous avons adoptée ici en montrant la convergence du processus $(Y_1, Y_2)^t$ va permettre de traiter des fonctionnelles de X plus complexes que ce dont nous avons besoin dans la section précédente. Les résultats relatifs au renouvellement (loi de $N_t e^{-\kappa\phi(t)}$, $H(m_{h_t}^{(N_t+1)})/t \dots$) dont nous avons parlé dans le paragraphe précédent deviennent des conséquences assez simples de cette convergence (en utilisant uniquement la seconde composante du processus $(Y_1, Y_2)^t$). Le supremum des temps locaux, quant à lui, va bien entendu utiliser les deux composantes.

La seconde étape est de montrer que la fonction de répartition $\mathbb{P}(\mathcal{L}_X^*(t)/t \leq \alpha)$ peut être encadrée précisément par des probabilités d'événements fonction exclusivement de $(Y_1, Y_2)^t$, on a en effet

Lemma 2.11. Soit $\alpha > 0$, pour tout t suffisamment grand,

$$\mathcal{P}_1^- - o(1) \leq \mathbb{P}(\mathcal{L}_X^*(t)/t \leq \alpha) \leq \mathcal{P}_1^+ + o(1), \text{ avec}$$

$$\mathcal{P}_1^\pm = \mathbb{P} \left[(1 - \bar{\mathcal{H}}_{\mathcal{N}_t-1}) \frac{\bar{\ell}_{\mathcal{N}_t} - \bar{\ell}_{\mathcal{N}_t-1}}{(\bar{\mathcal{H}}_{\mathcal{N}_t} - \bar{\mathcal{H}}_{\mathcal{N}_t-1})} \leq \alpha_t^\pm, \max_{1 \leq j \leq \mathcal{N}_t-1} (\bar{\ell}_j - \bar{\ell}_{j-1}) \leq \alpha_t^\pm \right],$$

et $\bar{\mathcal{H}}_k := Y_2^t(k e^{-\kappa\phi(t)})$, $\bar{\ell}_k := Y_1^t(k e^{-\kappa\phi(t)})$, $\alpha_t^\pm := \alpha(1 \pm (\log \log t)^{-1})$, $\mathcal{N}_t := \min\{m \geq 1, \bar{\mathcal{H}}_m \geq 1\}$.

On remarque que l'on retrouve dans ce Lemme exactement l'idée décrite après le Théorème 2.9. Il est obtenu à la fois en montrant que les temps locaux à l'extérieur des h_t -minima est négligeable, tout en intégrant le résultat du Lemme 2.10. Il faut également être suffisamment précautionneux pour garder toute l'information cruciale, faire apparaître la variable \mathcal{N}_t en est un exemple. Par le même calcul on peut aussi déduire une autre caractérisation de la loi limite du temps local, en effet on montre aussi que

$$\mathbb{P}(\mathcal{I} \leq \alpha) = \mathbb{E} \left[\int_0^{+\infty} a_\alpha(\mathcal{Y}_2(s)) \mathbb{1}_{\mathcal{Y}_1^\sharp(s) \leq \alpha, \mathcal{Y}_2(s) < 1} ds \right], \quad (5)$$

où pour tout $x \geq 0$, $a_\alpha(x) := \mathbb{E} \left[\left(\frac{\mathcal{R}_\kappa}{1-x} \right)^\kappa \mathbb{1}_{\frac{1-x}{\mathcal{R}_\kappa} \leq \alpha} \right]$.

L'interprétation de \mathcal{I} est un peu moins évidente mais fait apparaître explicitement la variable \mathcal{R}_κ : comme auparavant \mathcal{Y}_1^\sharp est le plus grand saut de \mathcal{Y}_1 avant que \mathcal{Y}_2 atteigne 1, c'est à dire le

supremum des temps locaux dans les vallées entièrement visitées. La fonction $a_\alpha(x)$ traduit ce qu'il se passe dans la dernière vallée où $1 - x$ est la proportion de temps restant à la particule lorsqu'elle atteint le dernier h_t -minima visité.

L'étape suivante est de passer à la limite en t pour \mathcal{P}_1^\pm en se servant de la convergence du processus $(Y_1, Y_2)^t$. Cela nécessite essentiellement l'étude de la continuité des applications qui a tout couple de fonctions (sous certaines hypothèses) (f_1, f_2) associe $f_1^\sharp(f_2(1))$ ou $f_1^\sharp(f_2(1)^-)$. On conclut ensuite par le théorème de continuité. Il se trouve que ce passage est plus facile à obtenir en utilisant la première caractérisation du supremum des temps locaux (via le Lemme 2.11) que la seconde qui permet d'obtenir (5). \square

La méthode utilisée permet en outre, d'obtenir directement d'autres fonctionnelles comme le supremum des temps locaux dans la dernière vallée franchie avant l'instant t et la suivante :

$$\frac{\mathcal{L}_X^*(H(m_{N_t}))}{t} \xrightarrow{\mathcal{L}} \mathcal{Y}_1^\sharp(\mathcal{Y}_2^{-1}(1)^-), \quad \frac{\mathcal{L}_X^*(H(m_{N_t+1}))}{t} \xrightarrow{\mathcal{L}} \mathcal{Y}_1^\sharp(\mathcal{Y}_2^{-1}(1)).$$

On voit que la première limite apparaît naturellement dans la loi \mathcal{I} , mais pas la seconde. Etant donné la forme explicite de la loi limite \mathcal{I} (en particulier (5)) on peut obtenir les queues de distribution suivantes : il existe $\lambda_0 > 0$ telle que

$$\log(\mathcal{I}([x, +\infty[)) \underset{x \rightarrow +\infty}{\sim} -8x, \quad \log(\mathcal{I}([0, x])) \underset{x \rightarrow +0}{\sim} \frac{\lambda_0}{x},$$

Pour comprendre ce qu'il se passe de différent entre les premières vallées et la dernière on a alors envie d'étudier la position du point favori, c'est un travail en cours qui clôturera cette étude.

2.4 Prolongements

- On s'attend à ce que les convergences en loi (pour le temps local) que nous avons obtenues soient très similaire pour le cas discret. La différence importante entre les deux venant de l'étude de l'équivalent discret de la variable \mathcal{R}_κ étudiée en détail dans les articles de Enriquez-Sabot-Zindy.

On aurait aussi envie de dire que tout problème relatif à ce processus (vu jusqu'à l'instant t) et dépendant fortement de ce qu'il se passe dans les h_t -minima peut être résolu en ajoutant une composante au processus Y_2^t comme nous l'avons fait pour le temps local. En particulier, le résultat de Tanaka peut être retrouvé en étudiant le processus $(Y_2^t(s), s > 0)$ associé au processus de drift pur obtenu comme limite de $(\sum_{i=1}^{\lfloor se^{\kappa\phi(t)} \rfloor} m_i - m_{i-1}/t^\kappa, s > 0)$.

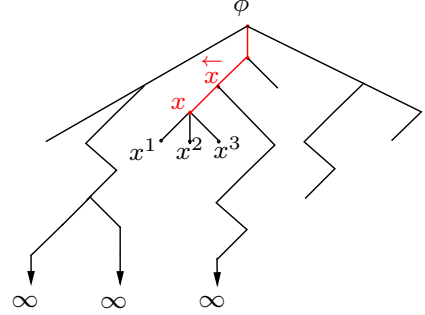
Il est maintenant naturel de s'interroger sur la convergence presque sûre de \mathcal{L}_X^* , d'obtenir les limites supérieures et inférieures et ainsi compléter ainsi les résultats de [GS02] pour le cas continu.

- Nous n'avons pas du tout parlé du cas $\kappa \geq 1$, en effet A. Devulder a étudié ce cas dans sa thèse [Dev06b], [Dev06a] ainsi que dans des travaux à venir.
- Pour cette diffusion unidimensionnelle continue, c'est clairement le cas où le milieu est Lévy (avec mesure de Lévy non triviale!) qui est le plus riche et qu'il faudrait maintenant privilégier. Rappelons quelques références ici [Car97], [Tan04], [Sin08], [Sin07a], [Sin07b]. G. Vechambre travaille actuellement sur l'extension des résultats de [10] et [15] en utilisant les hypothèses de [Sin08], à savoir un processus de Lévy spectralement négatif sans saut positif. Comme pour la diffusion en milieu continu, une étude fine des fonctionnelles exponentielles de ce processus de Lévy conditionné à rester positif ou négatif est nécessaire.
- Pour en revenir à la motivation initiale de la curiosité observée par Z. Shi, le temps local d'une diffusion est une densité qui donne une idée partielle du temps dépensé par le processus localement. On pourrait avoir envie, par exemple, de faire une étude de concentration comme on l'a fait dans le cas de la marche de Sinai pour le processus (et la marche) transient.

3 Temps discret, espace branchant discret

Dans ce paragraphe on étudie le cas de marches aléatoires sur un arbre de Galton-Watson avec conductances aléatoires.

Plus précisément, soit \mathbb{T} un arbre enraciné en ϕ appartenant à une famille d'arbres de Galton-Watson supercritique. Pour tout sommet x de \mathbb{T} on note N_x le nombre de ses descendants qui est donc une copie indépendante de la variable aléatoire \mathbf{N} à valeurs entières positives telle que $\mathbb{E}(\mathbf{N}) > 1$. x^1, x^2, \dots, x^{N_x} désigneront les sommets des enfants de x , et $\llbracket \phi, x \rrbracket$ l'ensemble des sommets du plus court chemin partant de ϕ menant à x . Aussi $\llbracket \phi, x \rrbracket := \llbracket \phi, x \rrbracket \setminus \{\phi\}$ et \overleftarrow{x} désignera l'ascendant de x . A chacun de ces sommets x on associe une variable aléatoire U_x à valeurs réelles. On définit maintenant la marche branchante V au sens de Biggins [Big77] que l'on appellera aussi potentiel :



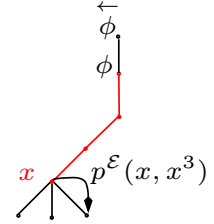
$$V(x) := \sum_{z \in \llbracket \phi, x \rrbracket} U_z, \quad z \in \mathbb{T} \setminus \{\phi\}, \quad V(\phi) = 0.$$

On désignera par $\mathcal{E} := (\mathbb{T}, V)$ un environnement fixé parmi les trajectoires des couples (arbres de G-W, marches branchantes).

Soit maintenant $(X_n, n \in \mathbb{N}, X_0 = \phi)$ une chaîne de Markov à valeurs dans \mathbb{T} , définie par les probabilités de transition :

$$p^{\mathcal{E}}(x, x^i) := \frac{e^{V(x) - V(x^i)}}{\sum_{j=1}^{N_x} e^{V(x) - V(x^j)} + 1}, \quad p^{\mathcal{E}}(x, \overleftarrow{x}) := 1 - \sum_{i=1}^{N_x} p^{\mathcal{E}}(x, x^i).$$

Pour éviter le problème en la racine, on utilise la même astuce que dans [FHS12] en ajoutant un parent à ϕ (et donc à \mathbb{T}) que l'on nomme bien entendu $\overleftarrow{\phi}$ et on pose $p(\overleftarrow{\phi}, \phi) = 1, V(\overleftarrow{\phi}) = 0$. (X_n, n) est donc une marche aléatoire au plus proche voisin dont on note $\mathbb{P}^{\mathcal{E}}$ la distribution à environnement fixé. Lorsque (X_n, n) est regardée sous la mesure de probabilité \mathbb{P} définie par $\mathbb{P}(A) = \int \mathbb{P}^{\mathcal{E}(\omega)}(A) dP(\omega)$, elle devient la marche aléatoire en milieu aléatoire que l'on étudie ici.



Ce type de marche appartient à la classe des marches aléatoires biaisées sur des arbres de Galton-Watson. Pour des U_x identiques constants ces processus ont été étudiés dans de nombreux travaux dont voici quelques références anciennes [Lyo90], [Lyo92], [LPP95], [LPP96] et plus récentes [PZ08], [AH12].

A partir de maintenant les variables aléatoires $\{U_x, x \in \mathbb{T}\}$ sont supposées indépendantes et de même loi et on note \mathbf{U} la variable générique ($U_x \stackrel{\mathcal{L}}{=} \mathbf{U}$). De plus on utilisera l'hypothèse d'uniforme ellipticité suivante : il existe $\mathbf{N}_0 > 0$ et $\mathbf{U}_0 > 0$ tels que presque sûrement $\mathbf{N} \leq \mathbf{N}_0$ et $|\mathbf{U}| \leq \mathbf{U}_0$. On peut alors définir pour tout $t > 0$ la transformée de log-Laplace

$$\psi(t) := \log \mathbb{E} \left(\sum_{|z|=1} e^{-tV(z)} \right),$$

où $|z|$ désigne la génération de z , ψ renferme l'essentiel des propriétés du milieu aléatoire.

Sur la Figure 2 on a représenté une première classification des marches (X_n, n) en fonction de ψ en terme de *classes récurrentes*. Ces critères ont été obtenus en plusieurs étapes en commençant par les travaux de [LP92] et [MP02] pour des arbres réguliers puis ceux de [Far11] dans notre contexte.

Nous ne parlerons pas du tout ici du cas transient qui est cependant un sujet très actif depuis un certain nombre d'années, en voici quelques références [AÖ8], [Aï0], [Aï1], [AFGH12], [AHOZ13] et deux thèses E. Aïdékon [AÖ9] et A. Fribergh [Fri09].

Pour les cas récurrents nous allons voir que suivant l'aspect étudié, on peut avoir des comportements identiques ou fortement différents. Il n'est pas inutile aussi d'avoir en tête l'allure du graphe de ψ dans chacun de ces cas (Figure 3). Les cas $\psi'(1) = 0$ et $\psi'(1) > 0$ sont en fait très semblables bien que l'un soit récurrent positif et l'autre récurrent nul. Aussi lorsque $\psi'(1) < 0$, apparaît l'existence d'une constante que nous appellerons κ qui est la racine de ψ strictement plus grande que 1 et possiblement infinie ($\kappa = \inf\{s > 1, \psi(s) = 0\}$).

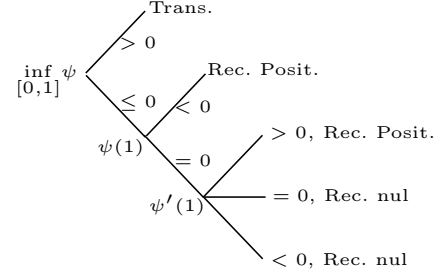


FIGURE 2 – Critères de récurrence

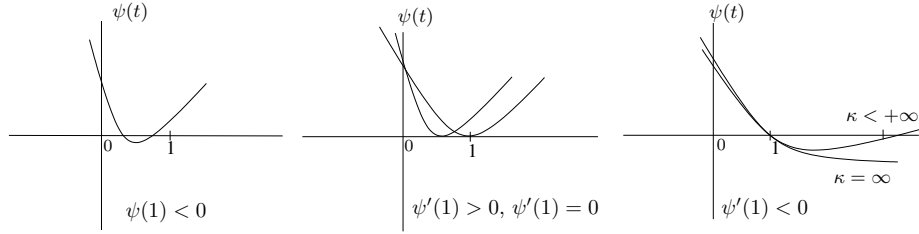


FIGURE 3 – ψ

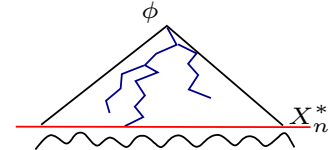
Dans le paragraphe suivant nous présentons sommairement les premiers travaux effectués dans le cas récurrent, ceci permettra de situer nos travaux.

3.1 Les travaux de Y. Hu et Z. Shi

Une première étude asymptotique sur la trajectoire $(X_k, k \leq n)$ est donnée par Y. Hu et Z. Shi, elle concerne la plus grande génération visitée notée X_n^* , définie par

$$X_n^* := \max_{0 \leq k \leq n} |X_k|.$$

Nous avons rassemblé les résultats de Hu-Shi datant de 2007 ([HS07b], [HS07a]) et ceux de Faraud-Hu-Shi [FHS12] de 2011 dans le Théorème qui suit. Cela mérite cependant quelques précisions, les résultats de 2007 sont obtenus pour des arbres réguliers ils couvrent tous les cas. Les résultats de 2011 sont obtenus pour des arbres de G-W sous $(\psi(1) = 0, \psi'(1) \geq 0)$ avec des hypothèses plus faible que l'uniforme ellipticité, en particulier sont précisées les constantes de convergence pour ces cas :



Théoreme 3.1 (Hu-Shi; Faraud-Hu-Shi). *\mathbb{P} -presque sûrement sur l'ensemble de non extinction de l'arbre, on a*

- Si $\psi(1) < 0$, $\lim_{n \rightarrow +\infty} X_n^* / \log n = a_1$,*
- Si $\psi(1) = 0, \psi'(1) > 0$, $\lim_{n \rightarrow +\infty} X_n^* / (\log n)^3 = a_2$,*
- Si $\psi(1) = 0, \psi'(1) = 0$, $\lim_{n \rightarrow +\infty} X_n^* / (\log n)^3 = a_3$,*

Si $\psi(1) = 0, \psi'(1) < 0, \lim_{n \rightarrow +\infty} \log X_n^* / \log n = 1 - 1/\min(\kappa, 2)$.
 a_1, a_2 et a_3 sont explicitement connues.

La grande richesse de comportements à l'intérieur de la seule classe des marches récurrentes est remarquable. Si on laisse de côté les deux cas récurrents positifs on se retrouve avec une marche très lente et une marche sous-diffusive/diffusive. Ainsi lorsque $\psi'(1) < 0$ le comportement rappelle un peu celui de la marche de Kesten-Kozlov-Spitzer [KKS75] cependant ici la marche est récurrente (remarquons que dans le cas transient sur l'arbre il existe aussi des cas non ballistiques [AÖ8]). Notons aussi que pour ce cas G. Faraud [Far11] montre que pour des κ grands > 5 , un théorème central limite fonctionnel peut être obtenu pour X^* .

Si $\psi'(1) = 0$ la marche est très lente et on a aussi envie de la rapprocher du comportement asymptotique de la marche de Sinai. Cependant il est à noter qu'ici la variable X_n^* peut avoir un comportement très différent de $|X_n|$. On va voir dans la Section 3.3 que plusieurs indices nous indiquent que cela devrait être effectivement le cas, les travaux récents de Hu-Shi [HS14] vont plus loin dans cette direction, en particulier, ils conjecturent pour $|X_n|$ que $|X_n|/(\log n)^2 \xrightarrow{\mathcal{L}} \mathcal{X}$. Je ne vais pas donner de détail de la preuve du Théorème 3.1 qui requiert de nombreux outils (décomposition spinale pour V , résultats de petites déviations de Mogulskii, inégalité de Paley-Zygmund ...) mais simplement donner une idée intuitive qui conduit à cette normalisation en $(\log n)^3$. On retient donc ici l'idée suivante, certes un peu caricaturale et ne justifiant que la borne inférieure mais qui donne une première idée : la marche, en un temps n et partant de la racine, acceptera de visiter sans difficulté tous sites de hauteur moindre que $\log n$ (on verra dans le paragraphe 3.3 que ces sites sont assez peu nombreux [voire (8)], et donc il est raisonnable de dire qu'ils seront tous visités avant l'instant n). La question est donc de savoir sur combien de génération le potentiel accepte d'avoir des fluctuations maximales plus petites que $\log n$, Fang-Zeitouni [FZ] et Faraud-Hu-Shi [FHS12] répondent (indépendamment) à cette question de la façon suivante : \mathbb{P} -presque sûrement sur l'ensemble de non extinction

$$\lim_{m \rightarrow +\infty} \frac{\min_{|z|=m} \bar{V}(z)}{m^{1/3}} = a, \text{ avec } \bar{V}(z) := \max_{u \in \llbracket \phi, z \rrbracket} V(u).$$

Si on prend $m^{1/3} = \log n$, on obtient donc cette normalisation en $(\log n)^3$. On retiendra aussi que la variable clé, dépendant uniquement de l'environnement, dans l'étude de X_n^* est ainsi $\min_{|z|=m} \bar{V}(z)$.

▷ Un autre aspect qui peut être intéressant est de comprendre comment l'arbre est visité par la marche, c'est à dire essayer de décrire le nuage de points visités jusqu'à un instant n . On va commencer par décrire la plus grande génération qui est entièrement visitée avant l'instant n . J'ai travaillé sur ce point avec P. Debs. ▷

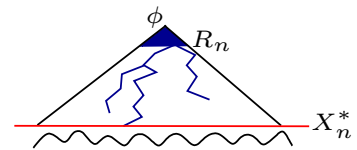
3.2 La plus grande génération entièrement visitée

Définissons la variable qui va nous intéresser dans cette section :

$$R_n := \max \left\{ k, \min_{|z|=k} \mathcal{L}(z, n) \geq 1 \right\} \text{ où } \mathcal{L}(z, n) := \sum_{i=1}^n \mathbb{1}_{X_i = z}.$$

Pour établir la convergence de cette variable on a besoin de la constante qui apparaît dans la loi des grands nombres des marches branchantes [Big76].

$$J(a) := \inf_{t \geq 0} \{\psi(-t) - at\}, \gamma := \sup\{a \in \mathbb{R}, J(a) > 0\}.$$



On a alors le résultat,

Théoreme 3.2 (A.-Debs [8]). \mathbb{P} -presque sûrement sur l'ensemble de non extinction de l'arbre, si $\psi'(1) < 0$, $\lim_{n \rightarrow +\infty} R_n / \log n = 1/\gamma \times 1/\min(\kappa, 2)$, sinon $\lim_{n \rightarrow +\infty} R_n / \log n = 1/\gamma$.

On voit immédiatement que le comportement de R_n est moins riche que celui de X_n^* . On peut même aller plus loin dans la similarité des 4 cas récurrents, il suffit pour cela de regarder la variable R jusqu'à n retours en la racine : définissons $T_x^n = \inf\{k > T_x^{n-1}, X_k = x\}$ pour $n \geq 1$ et $T_x^0 = 0$, on a pour tous les cas : \mathbb{P} -presque sûrement sur l'ensemble de non extinction

$$\lim_{n \rightarrow +\infty} R_{T_\phi^n} / \log n = 1/\gamma. \quad (6)$$

Nous passons maintenant aux idées de preuve, qui donnent aussi l'idée intuitive du résultat.

Idées de preuve :

On commence naturellement par étudier $R_{T_\phi^n}$.

La borne inférieure est obtenue simplement avec un raisonnement similaire à celui que l'on a l'habitude de faire dans le cas uni-dimensionnel : à la génération $r_- := \frac{\log n}{\gamma}(1 - \epsilon)$ presque sûrement pour tout n grand $\max_{|z|=r_-} \bar{V}(z) \leq (1 - \epsilon/2) \log n$. On montre alors que presque sûrement en n retours en la racine la marche visite tous les points de la génération r_- .

Pour la borne supérieure, il faut montrer que la marche n'accepte pas de visiter tous les points d'une génération $(1 + \epsilon) \log n / \gamma$. On sait qu'à cette génération au moins un point sera tel que \bar{V} sera plus grand que $(1 + \epsilon/2) \log n$, cependant pour montrer la convergence presque sûre ceci n'est pas suffisant. Il faut aller un peu plus loin et remarquer que presque sûrement on peut trouver un nombre de sites aussi grand que l'on veut (comme fonction de ϵ), tel que pour ces sites \bar{V} est plus grand que $(1 + \epsilon/2) \log n$. Ces sites auront de plus leur ancêtre commun proche de la racine à une génération dépendant uniquement de ϵ . Si on veut que ces sites soient visités, cela va obliger la marche à atteindre des niveaux de potentiels élevés en de nombreux points très distants les uns des autres. Très distants au sens où leur ancêtre commun est proche de la racine. Cette grande distance entre ces individus va permettre de découper la trajectoire de la marche en parties indépendantes. Ainsi on obtient à la fois des événements rares et indépendants qui se répètent, cela permet d'améliorer la convergence de la probabilité qui permet ainsi d'appliquer Borel-Cantelli.

Retour à R_n , on jouit ici du fait que la connaissance du comportement logarithmique du temps local en la racine est suffisant. En utilisant les résultats de Hu-Shi [HS07b] et Faraud-Hu-Shi [FHS12], on montre que \mathbb{P} .p.s sur l'ensemble de non extinction

$$\begin{aligned} \text{Si } \psi'(1) < 0, \quad \lim_{n \rightarrow +\infty} \frac{\log \mathcal{L}(\phi, n)}{\log n} &= \frac{1}{\min\{\kappa, 2\}}, \\ \text{sinon } \lim_{n \rightarrow +\infty} \frac{\log \mathcal{L}(\phi, n)}{\log n} &= 1. \end{aligned}$$

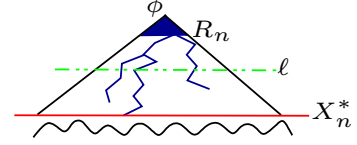
Cette dernière remarque associée à (6) implique le résultat, de plus on voit que la variable aléatoire clé est ici la variable bien connue $\max_{|z|=m} \bar{V}(z)$ [en particulier des résultats de C. McDiarmid [McD95] sont utilisés pour la preuve] dont le comportement asymptotique ne dépend pas des hypothèses de récurrence particulières (d'où le comportement universel pour $R_{T_\phi^n}$). \square

\triangleright Au delà de la génération R_n , le nombre de sites visités va rapidement devenir négligeable comparativement au nombre de sites sur une génération. Ainsi dès la génération $(1 + \epsilon)(\log n)/\psi(0)$,

le nombre de sites visités sera très grand devant le temps imparti à la marche n . D'un point de vu quantitatif on peut donc se poser la question suivante, prenons une génération (déterministe) entre R_n et X_n^* , quel est le nombre de points visités sur cette génération ? \triangleright

3.3 Nombre de sites visités à une génération donnée

Je parlerai ici uniquement du cas lent $\psi(1) = \psi'(1) = 0$, et je supposerai que V est non-arithmétique. Dans la Section Prolongements je discuterai sommairement du cas $\psi'(1) < 0$ pour comparaison. Les résultats de cette section ont été obtenus avec P. Debs [9]. Nous nous intéressons donc aux variables aléatoires suivantes : soit $\ell > R_n$ un entier,



$$M_n := \sum_{|z|=\ell} \mathbb{1}_{\mathcal{L}(z,n) \geq 1}, \quad K_n := M_{T_n^\phi}.$$

Le problème s'avère plus difficile que l'étude de R_n , nous nous intéressons donc à la moyenne de K_n . Un outil très utile ici [qui est aussi utilisé dans la majorité des articles que j'ai cités sur ce thème y compris la section précédente], est l'identité de Biggins-Kyprianou souvent appelée Many to One Lemma (voir [Shi11] et [BK97]). Je la rappelle donc ci-dessous dans le cas général : pour tout $n \geq 1$ et toute fonction Borélienne $F : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, +\infty)$,

$$\mathbb{E} \left[\sum_{|x|=n} e^{-V(x)-\psi(1)n} F(V(x), 1 \leq i \leq n) \right] = \mathbb{E}_Q[F(S_i, 1 \leq i \leq n)]$$

où $(S_i - S_{i-1})_{i \geq 1}$, sont des variables aléatoires i.i.d sous la mesure Q (obtenue par le théorème d'extension de Kolomogorov à partir de \mathbb{P} et de la martingale $\sum_{|z|=n} e^{-V(z)-\psi(1)n}$), telles que

$$\mathbb{E}_Q[G(S_1)] = \mathbb{E} \left[\sum_{|x|=1} e^{-V(x)-\psi(1)} G(V(x)) \right],$$

pour toute fonction Borélienne $G : \mathbb{R} \rightarrow [0, +\infty)$. x_i désigne ici le $i^{\text{ème}}$ descendant de ϕ du plus court chemin de ϕ à x .

Etant donné qu'ici $\psi(1) = 0$, on a l'identité $e^{\psi(1-x)} = \mathbb{E}_Q(e^{xS_1})$. Notons f , la fonction suivante définie pour tout x suffisamment petit

$$f(x) := 1 - \frac{x}{2\sigma^2} + x^2\lambda(x),$$

où λ est la série de Cramér relative à S_1 ([Pet75] Chapitre VIII, section 2). A partir de maintenant, on supposera que

$$\ell := (\log n)^{1+\zeta}, \quad 0 < \zeta < 2,$$

on obtient le résultat suivant :

Théoreme 3.3 (A.-Debs [9]). *Pour tout $0 < \zeta < 2$, $\epsilon > 0$ ne dépendant pas de ζ , il existe deux constantes positives C_1 et C_2 telles que pour tout n suffisamment grand*

$$\frac{C_1}{(\log n)^\epsilon} \frac{e^{(\log n) \cdot f[(\log n)^{-\zeta}]}}{(\log n)^{(1+\tilde{\zeta})/2}} \leq \mathbb{E}[K_n(\ell)] \leq C_2 \frac{e^{(\log n) \cdot f[(\log n)^{-\zeta}]}}{(\log n)^{(1+\tilde{\zeta})/2}},$$

où $\tilde{\zeta} := \mathbb{1}_{0 < \zeta < 1} + \zeta \mathbb{1}_{1 \leq \zeta < 2}$.

On peut voir que la moyenne de $K_n(\ell)$, présente une transition de phase en $\zeta = 1$. En effet par définition de f ,

$$(\log n)f[(\log n)^{-\zeta}] = \log n - \frac{(\log n)^{1-\zeta}}{2\sigma^2} + (\log n)^{1-2\zeta}\lambda((\log n)^{-\zeta}).$$

Ainsi pour tout $0 < \zeta < \zeta' \leq 1$,

$$\lim_{n \rightarrow +\infty} \frac{\mathbb{E}[K_n(\ell')]}{\mathbb{E}[K_n(\ell)]} = +\infty, \text{ avec } \ell' = (\log n)^{1+\zeta'}$$

alors que si $1 \leq \zeta < \zeta' < 2$,

$$\lim_{n \rightarrow +\infty} \frac{\mathbb{E}[K_n(\ell')]}{\mathbb{E}[K_n(\ell)]} = 0.$$

Ceci nous incite à dire que les générations de l'ordre de $\ell_c := (\log n)^2$ sont en moyenne les générations les plus visitées en terme de sites visités distincts.

On remarque aussi que pour $1/2 < \zeta < 1$, la correction dans f venant de la série de Cramér disparaît, et que si $1 \leq \zeta < 2$, l'exponentiel de $(\log n) \cdot f[(\log n)^{-\zeta}]$ se réduit à $\text{const} \cdot n$. On va maintenant donner quelques idées de la preuve qui n'est malheureusement pas très intuitive, cependant nous verrons que cela donne des indications sur la voie à suivre pour obtenir une convergence en probabilité.

Idée de preuve

On calcule successivement la moyenne sous $\mathbb{P}^{\mathcal{E}}$ de K_n puis la moyenne de cette moyenne. $\mathbb{E}^{\mathcal{E}}(K_n)$ peut s'écrire essentiellement comme combinaison linéaire des deux variables aléatoires :

$$\sum_{|z|=\ell} \mathbb{1}_{\bar{V}(z) \leq \log n}, \text{ et } \sum_{|z|=\ell} np_z \mathbb{1}_{\bar{V}(z) > \log n}, \quad (7)$$

où $p_z := (\sum_{u \in \llbracket \phi, z \rrbracket} e^{V(u)})^{-1}$.

On remarque que la première somme compte les sites dont le potentiel est (toujours) plus petit que $\log n$ jusqu'à la génération ℓ . La seconde tient compte des sites z dont le potentiel est plus haut que $\log n$ pondéré par np_z . L'absence de p_z dans le premier terme fait que sa moyenne est plus simple à calculer. En utilisant le MOL, ainsi que des résultats de déviation pour des sommes de variables i.i.d. avec condition de Cramér on obtient

$$\mathbb{E}\left(\sum_{|z|=\ell} \mathbb{1}_{\bar{V}(z) \leq \log n}\right) \asymp \left(\frac{\mathbb{1}_{0 < \zeta < 1}}{(\log n)^{1+\zeta}} + \frac{\log n}{(\log n)^{3(1+\zeta)/2}} \mathbb{1}_{1 \leq \zeta < 2}\right) e^{\log n f(\log n^{-\zeta})}. \quad (8)$$

On voit alors immédiatement par rapport au résultat du théorème que la contribution des sites tels que V reste en dessous de $\log n$ est négligeable. Cela montre que les points de potentiels élevés (plus élevés que $\log n$) sont importants et même fondamentaux ce qui contraste avec la MAMA unidimensionnelle. Remarquons ici qu'un travail récent de Hu-Shi [HS14] montre qu'en fait la marche peut visiter des hauteurs de potentiel de l'ordre de $(\log n)^2$ presque sûrement ! Pour calculer la moyenne du second terme il faut travailler un peu plus à cause du p_z après le MOL on obtient qu'il faut calculer $E\left((\sum_{i=1}^{\ell} e^{S_i - S_{\ell}})^{-1} \mathbb{1}_{\bar{S}_{\ell} > \log n}\right)$, la borne supérieure ne pose pas de problème car l'inverse de la somme est inférieure à $e^{S_{\ell} - \bar{S}_{\ell}}$ ($\bar{S}_{\ell} := \max_{1 \leq i \leq \ell} S_i$) et on applique les mêmes techniques que pour la partie négligeable. Pour la borne inférieure on doit à la fois maîtriser les fluctuations de la somme $\sum_{i=1}^{\ell} e^{S_i - S_{\ell}}$ tout en gérant la dépendance avec l'événement dans l'indicatrice. On partitionne donc suivant les valeurs du temps d'atteinte du

maximum de S avant l'instant ℓ , cela conduit à ce que la somme $\sum_{i=1}^{\ell} e^{S_i - S_\ell}$ s'écrive comme la somme de deux morceaux indépendants. On montre ensuite que ces deux sommes ne sont pas très grandes (c'est à dire plus petites que $\delta \ell^\epsilon$, pour $0 < \delta < 1/2$). Le cas le plus délicat est celui qui correspond à la trajectoire de S avant l'atteinte du maximum car c'est là où la dépendance avec l'événement dans l'indicatrice est réalisée. Remarquons que nous avons déjà eu à faire à des sommes d'exponentielles de différences de potentiels assez similaires dans la Partie 1 (voire (1) par exemple), et que l'apparition de ces objets ici vient du fait que les probabilités du type $\mathbb{P}_\phi^\mathcal{E}(T_z < T_\phi)$ s'écrivent comme dans le cas unidimensionnel, en ne tenant compte que de V sur le chemin de ϕ à z , i.e. $\mathbb{P}_\phi^\mathcal{E}(T_z < T_\phi) \sim p_z$.

Enfin la variable clé ici est donc la seconde variable aléatoire dans (7). \square

▷ Jusqu'ici nous nous sommes intéressés à l'aspect quantitatif, on voit que les résultats ne sont pas encore très précis, en particulier on aimerait avoir au moins un résultat en probabilité sur K_n et M_n . Je discuterai de petites avancées récentes dans la partie Prolongements. Dans la section suivante nous donnons une première idée de la façon dont les points visités se répartissent sur l'arbre. \triangleright

3.4 Etalement des points visités

Pour ce paragraphe on se place encore sous l'hypothèse $\psi(1) = \psi'(1) = 0$. On aura besoin à la fois d'identifier des ensembles de sites contigus compris entre deux mêmes générations, ainsi que d'une notion de distance entre de tels ensembles.

Introduisons donc d'abord ce que nous appellerons *cluster*. Soit $z \in \mathbb{T}$ et $m > |z|$, l'unique cluster issu de z et de génération m que l'on note $\mathcal{C}_m(z)$ est l'ensemble des descendants u de z tels que $|u| = m$, soit

$$\mathcal{C}_m(z) := \{u > z, |u| = m\}.$$

$u > z$ signifiant donc que u est un descendant de z . Par convention on supposera que $\mathcal{C}_{|z|}(z)$ est vide, sinon remarquons qu'un cluster $\mathcal{C}_m(z)$ n'est vide que si z n'a pas de descendant à la génération m .

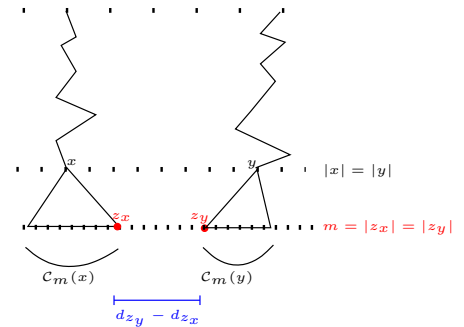
Si on se donne deux clusters $\mathcal{C}_m(z)$ et $\mathcal{C}_m(z')$ avec $z \neq z'$ et $|z| = |z'|$, on a besoin de définir le nombre d'individus entre eux. Pour cela commençons par rappeler la notation de Neveu [Nev86] introduisant un ordre partiel sur l'arbre. A tout sommet x de génération $m \in \mathbb{N}$ on associe la suite $x_1 \dots x_m$ ou $x_i \in \mathbb{N}$, nous écrirons donc $x = x_1 \dots x_m$.

Cette suite donne la généalogie complète de x : si $y = x_1 \dots x_i$ avec $|y| = i < m$, alors y est l'unique ancêtre de x de la génération i et nous notons $y < x$. Ainsi par exemple $\overleftarrow{x} = x_1 \dots x_{m-1}$ et $1 \leq x_m \leq N_{\overleftarrow{x}}$, et x est l'unique x_m ème descendant de \overleftarrow{x} .

Pour étendre cet ordre partiel à des individus de la même génération $|x| = |z|$, on écrit $x < z$ s'il existe $i < m$ tel que $x_k = z_k$ pour $k < i$ et $x_i < z_i$.

On peut donc numéroté les individus d'une génération m de la gauche vers la droite, et on associe ainsi à chaque individu u un nombre d_u , qui est compris entre 1 (pour l'individu le plus à gauche) et $\text{card}[\mathcal{C}_m(\phi)]$ (pour l'individu le plus à droite). Maintenant, pour tout

A sous-ensemble de $\{z \in \mathbb{T}, |z| = m\}$, on note $\inf A := \inf\{d_z, z \in A\}$ (respectivement $\sup A := \sup\{d_z, z \in A\}$) le minimum dans A (respectivement le maximum dans A) asso-



cié à cette numérotation.

Proposition 3.4. *Soient $0 < \zeta < 2$ et $\epsilon > 0$ rappelons que $\ell = (\log n)^{1+\zeta}$*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\min_{z \in \mathcal{C}_{\ell^{1/3}}(\phi)} \max_{y > z, |y| = \ell} \mathcal{L}(y, n) \geq 1 \right) = 1.$$

Cette proposition montre simplement que tous les individus d'une génération assez proche de la racine $\ell^{1/3}$ ont au moins un descendant visité à la génération ℓ (Figure 4). Conservons maintenant un individu sur deux (en allant de gauche à droite) de cette génération $\ell^{1/3}$, on sait (avec une probabilité proche de 1) qu'à cette génération il y a au moins $e^{\epsilon(1-\epsilon)\psi(0)\ell^{1/3}/2}$ tels individus (z_1, z_3, \dots). De plus (toujours avec une probabilité proche de 1) on a aussi $\min_{j \text{ pair}} \text{card}[\mathcal{C}_\ell(z_j)] \geq e^{\psi(0)\ell(1-\epsilon)/2}$, on en déduit qu'au moins $e^{\epsilon(1-\epsilon)\psi(0)\ell^{1/3}/2}$ individus de génération ℓ séparés par $e^{\psi(0)\ell(1-\epsilon)/2}$ individus de la même génération ℓ , sont visités. On peut en conclure qu'il y a une certaine diffusion des sites visités sur l'arbre. La borne inférieure de $\min_{j \text{ pair}} \text{card}[\mathcal{C}_\ell(z_j)]$ est obtenue à partir de la queue de distribution à gauche des $\text{card}[\mathcal{C}_\ell(z_j)]$. Comme nous n'avons pas besoin d'une grande précision nous n'avons pas différencié les cas Schröder et Böttcher, la queue Schröder étant suffisante.

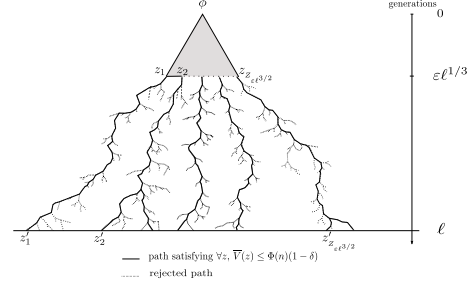


FIGURE 4 – Sites visités distants

Nous regardons maintenant si les points visités peuvent être regroupés :

Proposition 3.5. *Soit $\epsilon > 0$, pour n suffisamment grand notons $w_n := (1 - \epsilon) \log n / \gamma$, on a*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\max_{|z| = \ell - w_n} \min_{y \in \mathcal{C}_\ell(z)} \mathcal{L}(y, n) \geq 1 \right) = 1.$$

Ce résultat implique l'existence d'un cluster dont l'origine est la génération $\ell - w_n$ et la fin la génération ℓ entièrement visitée (Figure 5). Ce cluster entièrement visité a un nombre d'individus au moins égal à $e^{\psi(0)(1-2\epsilon)\log n/\gamma}$, c'est à dire du même ordre que le nombre d'individus de la génération R_n . Ce résultat doit d'ailleurs être aussi vrai presque sûrement car on le montre en utilisant les résultats de Faraud-Hu-Shi sur le $\min \bar{V}$ et les idées de la Section 3.2 sur R_n . Cela implique aussi une borne inférieure grossière pour $M_n(\ell)$ car la taille de ce cluster entièrement visité est une borne inférieure pour $M_n(\ell)$. Nous ne donnons pas d'idée de preuve de ces deux propositions car cela utilise les mêmes arguments que la preuve du Théorème suivant dont nous donnerons les idées.

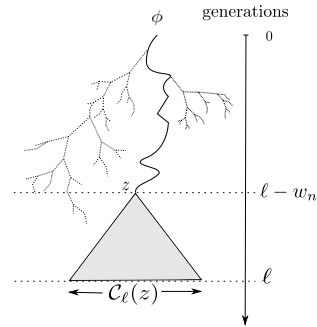


FIGURE 5 – Grand cluster visité

Le prochain résultat est un peu un mélange des deux propositions précédentes. Il montre que de petits clusters visités très distants les uns des autres et n'appartenant pas à un même grand cluster entièrement visité existent avec une grande probabilité.

Pour cela nous avons besoin de nouvelles notations, on se donne donc deux générations initiale et finale et \mathcal{C} un ensemble de clusters disjoints définis par ces générations. On note $(\mathcal{D}_j, 1 \leq j \leq |\mathcal{C}|)$, la suite ordonnée des clusters de \mathcal{C} , c'est à dire tels que pour tout j $\sup \mathcal{D}_j < \inf \mathcal{D}_{j+1}$. Définissons la distance³ minimale entre clusters de \mathcal{C} de la façon suivante

$$\mathbf{D}(\mathcal{C}) := \min_{1 \leq j \leq |\mathcal{C}|-2} (\inf \mathcal{D}_{j+2} - \sup \mathcal{D}_j),$$

La raison pour laquelle on saute un cluster dans la définition de \mathbf{D} vient du fait que l'on cherche des clusters très séparés les uns des autres. En évitant un cluster dans la définition on s'assure ainsi qu'il y a un minimum d'individus séparant deux clusters qui nous intéressent.

Théoreme 3.6 (A.-Debs [9]). *Soient $(k_n, h_n, r_n, n > 0)$ une suite composée d'entiers positifs tels que $k_n r_n + (k_n - 1)h_n = \ell$ pour tout n . Pour tout $1 \leq i \leq k_n$, notons \mathcal{C}_i un ensemble de clusters initiés à la génération $(i - 1)(r_n + h_n)$ et de points terminaux à la génération $ir_n + (i - 1)h_n$ (voir Figure 6), définissons l'événement suivant pour tout $m > 0$ et $q > 0$*

$$\mathcal{A}_i(m, q) := \bigcup_{\mathcal{C}_i} \left\{ \{ \text{card}(\mathcal{C}_i) \geq q, \mathbf{D}(\mathcal{C}_i) \geq m \} \cap \bigcap_{\mathcal{D} \in \mathcal{C}_i} \{ \forall z \in \mathcal{D}, \mathcal{L}(z, n) \geq 1 \} \right\}.$$

Il existe $0 < \mathbf{k} < 1 \wedge \zeta$, $0 < \mathbf{r} < 1$ avec $0 < \mathbf{k} + \mathbf{r} \leq 1$ tels que pour $k_n = (\log n)^{\mathbf{k}}$, $r_n = (\log n)^{\mathbf{r}}$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\bigcap_{i=2}^{k_n} \mathcal{A}_i(e^{\psi(0)h_n/2}, e^{\psi(0)r_n(i-1)/2}) \right) = 1. \quad (9)$$

Ce résultat n'est pas très lisible mais son écriture est difficilement simplifiable. De plus nous avons fait un abus de notation en ce qui concerne l'événement dans la probabilité en (9). En effet les clusters descendent les uns des autres ce qui n'apparaît pas, il faudrait plutôt écrire l'événement sous la forme $\{\dots \cap \{\dots \cap \{\dots\}\}\}$. La Figure 6 aide, je l'espère, à rendre cet abus tolérable. L'idée est de montrer que l'on peut trouver régulièrement dans l'arbre des clusters entièrement visités disjoints descendants les uns des autres. Le résultat dit que pour tout $i \leq k_n$, aux générations de la forme $ir_n + (i - 1)h_n$, on peut trouver $e^{\psi(0)r_n(i-1)/2}$ clusters de taille au moins $e^{\psi(0)r_n/2}$ entièrement visités. De plus la distance entre deux de ces clusters (au sens défini précédemment) est d'au moins $e^{\psi(0)h_n/2}$ individus. On en déduit alors par la décomposition de $\ell = k_n r_n + (k_n - 1)h_n$ que $h_n \sim (\log n)^{1+\zeta-\mathbf{k}}$ et du fait que $\mathbf{k} < \zeta$ que $e^{\psi(0)h_n/2} > n$. Ceci implique donc qu'entre deux clusters visités il y a au moins n sites et donc au moins 1 n'est pas visité puisque le temps imparti à la marche est n . Ces petits clusters visités ne font donc pas partie d'un même grand cluster entièrement visité.

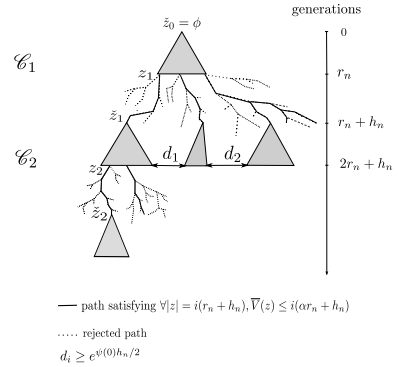


FIGURE 6 – Clusters visités

Idées de preuve : Elle s'appuie sur un découpage régulier dans l'arbre tel que représenté dans la Figure 6. Dans ce découpage on alterne des coupes franches entre les générations $g_i^1 := ir_n + ih_n$ et $g_i^2 := (i + 1)r_n + ih_n$ où l'ensemble des individus sont conservés, et des coupes sélectives

3. \mathbf{D} n'est pas une distance au sens usuel.

entre les générations g_i^2 et g_{i+1}^1 où un unique individu est conservé. Lors des coupes franches on laisse le potentiel (recentré) augmenter à son maximum c'est à dire dans le pire des cas $\mathbf{U}_0 r_n$. Lors des coupes sélectives, on conserve uniquement les descendants (de génération g_{i+1}^1) des individus des coupes franches tels que le potentiel (recentré) a le moins fluctué. Par le résultat de Faraud-Hu-Shi [FHS12] on sait que la plus petite fluctuation que l'on peut obtenir entre les générations g_i^2 et g_{i+1}^1 est de l'ordre de $(g_{i+1}^1 - g_i^2)^{1/3} = h_n^{1/3}$. Ainsi à la génération ℓ qui nous intéresse la hauteur maximale du potentiel atteinte pour les individus sélectionnés est au plus de l'ordre de $H_n := k_n * r_n \alpha + k_n * h_n^{1/3}$. On choisit alors k_n , r_n et h_n de manière à ce que H_n soit plus petit que $(1 - \epsilon) \log n$ (en fait des contraintes pour obtenir la convergence des probabilités font qu'au final H_n est plutôt de l'ordre de $(\log n)^{1-\epsilon}$). On sait alors que tous ces individus sont visités avec une grande probabilité, rappelons en effet qu'ils sont peu nombreux de manière triviale. On obtient aussi que $h_n \sim (\log n)^{1+\zeta-\mathbf{k}}$, ce qui implique le fait que les clusters sont disjoints. \square

3.5 Prolongements

• **Cas** $\psi(1) = 0$, $\psi'(1) = 0$

Rappelons que dans ce cas spécifique, on a noté $\ell_c = (\log n)^2$, avec X. Chen, nous avons un peu amélioré les résultats en montrant

Proposition 3.7 (A. -Chen).

$$\lim_{n \rightarrow +\infty} \frac{\log n}{n} \mathbb{E}(K_n(\ell_c)) = C \in (0, +\infty),$$

et pour tout $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{\log[K_n(\ell_c)]}{\log n} - 1 \right| \leq \epsilon \right) = 1,$$

le résultat ci-dessus étant aussi vrai si on remplace $K_n(\ell_c)$ par $M_n(\ell_c)$.

Ceci n'est pas suffisant pour montrer que la génération critique, c'est à dire la plus visitée, est $(\log n)^2$, cependant on peut naturellement penser que l'on a

$$M_n(\ell_c) \sim C \frac{n}{(\log n)^{\alpha+1}} \quad (10)$$

en probabilité, si le temps local en la racine se comporte en $\frac{n}{\log n}^4$, α n'est pas encore connu ! Pour résoudre ce problème le calcul de la moyenne de la variance quenched de $K_n(\ell_c)$ (pour des sites de potentiels hauts) révèle deux scénarios possibles. Un de ces scénarios indique que ce serait au niveau des générations $(\log n)^2$ qu'une fluctuation atypique du potentiel (que l'on sait caractériser) a la plus grande chance d'apparaître. Cette fluctuation atypique expliquerait la raison pour laquelle la génération critique pour le nombre de points distincts visités et peut-être aussi la localisation de $|X_n|$ se ferait au niveau des générations de l'ordre de $(\log n)^2$. Ceci nous fait bien entendu penser à l'idée intuitive que Sinai donne dans [Sin82b] (à propos de la fluctuation localisante du potentiel) que nous avons rappelé dans l'introduction pour expliquer le comportement de la MAMA unidimensionnelle récurrente.

Une fois ce résultat confirmé, il est naturel de se poser la question de la convergence presque sûre pour (M_n, n) , en particulier l'existence (ou pas) de la limite. On note d'ailleurs pour l'instant que contrairement au cas unidimensionnel (marche de Sinai par exemple), les limites supérieure

4. Ceci semble être le cas d'après des travaux à venir de Y. Hu et Z. Shi.

et inférieure des objets (X_n^* ou $R_n \dots$) que nous avons présentées ici ont toujours coïncidé (pour la MAMA uniquement car pour le milieu il y a des différences [HS09]).

Il y a aussi la description du nuage des points visités qui m'intéresse particulièrement. Encore une fois les premiers résultats que nous avons obtenus ne sont pas très précis, l'idéal serait comme dans le cas unidimensionnel d'avoir une description quenched des points visités.

Une fois ces questions résolues on pourrait s'attarder sur la génération (suffisamment distante de la racine) la plus visitée mais en terme de temps local (qui était ma motivation première, comme prolongement des idées de [6]). Ce n'est, a priori, pas évident que cette génération soit identique aux problèmes précédents.

• **Cas** $\psi(1) = 0$, $\psi'(1) < 0$

Le calcul de la moyenne de $K_n(\ell)$ est ici simplifiée par le fait que via le MOL la variable S_1 est maintenant de moyenne $|\psi'(1)|$, on a montré avec P. Debs que la génération critique est $\ell_c = \frac{\log n}{|\psi'(1)|}$, on obtient que pour tout $\epsilon > 0$

$$\begin{aligned} \forall \ell \geq \ell_c(1 + \epsilon), \quad \lim_{n \rightarrow +\infty} \frac{\mathbb{E}[K_n(\ell)]}{n} &= C' > 0, \\ \forall \ell \leq \ell_c(1 - \epsilon), \quad \lim_{n \rightarrow +\infty} \frac{\mathbb{E}[K_n(\ell)]}{n} &= 0. \end{aligned}$$

On peut être étonné par ces convergences notamment lorsque $\ell \geq \ell_c(1 + \epsilon)$, on s'attendrait en effet, puisque la marche est plus rapide, que moins de sites soient visités que dans le cas $\psi'(1) = 0$. Cependant il ne faut pas oublier que le temps local en ϕ en un temps n de cette marche est de l'ordre de $n^{\frac{1}{\min(\kappa, 2)}}$, on s'attend donc plutôt au résultat suivant : pour tout $\ell_c(1 + \epsilon) \leq \ell \leq n^{1 - (1/\min(2, \kappa)) - \epsilon}$ en probabilité,

$$M_n(\ell) \sim C n^{\frac{1}{\min(\kappa, 2)}}.$$

Ce qui est bien plus petit que (10).

4 Applications

Dans cette partie on utilise le modèle unidimensionnel de la Partie 1. On reprend donc, sauf mention contraire, l'essentiel des notations de celle-ci.

4.1 Estimation pour la marche de Sinai

Le travail [7] est une application des résultats obtenus dans le premier article de concentration ([2]). L'idée est d'estimer le potentiel \mathbf{S} à partir de la connaissance du temps local. Comme on a pu le voir une coordonnée importante pour la marche est $m_{\log n}$. Ce point peut facilement être approché par le point favori de la marche par le résultat suivant : soit $n > 1$ il existe un sous ensemble d'environnements \mathcal{E}_n et un entier n_0 tel que pour tout $n > n_0$, $\mathbb{P}[\mathcal{E}_n] \geq 1 - \log_2 n / \log n$ et

$$\inf_{\omega \in \mathcal{E}_n} \mathbb{P}^\omega \left[\max_{x \in \mathbb{F}_n} |m_{\log n} - x| \leq (\log_2 n)^2 \right] \geq 1 - (\log n)^{-2}.$$

où $\mathbb{F}_n := \{k \in \mathbb{Z}, \mathcal{L}(k, n) = \mathcal{L}^*(n)\}$.

En étendant le résultat de concentration [2] on va voir que l'on peut estimer à l'aide du temps local les différences de potentiel entre les points suffisamment visités et $m_{\log n}$. Soit $\gamma > 6$, notons \mathbb{L}_n^γ la zone d'estimation

$$\mathbb{L}_n^\gamma = \left\{ l \in \mathbb{Z}, \sum_{j=T_{k_n^*}}^n \mathbb{1}_{X_j=l} \geq ((\log n)^\gamma) \vee (\mathcal{L}(l, T_{k_n^*})) \right\}.$$

où $k_n^* = \inf\{|k|, k \in \mathbb{F}_n\}$ et rappelons que $T_x = \inf\{k \in \mathbb{N}^*, X_k = x\}$. Notons $\{S_k^n := (S_k - S_{m_{\log n}}) / \log n, k \in \mathbb{Z}\}$, les différences que l'on cherche à estimer et $(\hat{S}_l^n := 1 - \log \mathcal{L}(l, n) / \log n, l \in \mathbb{Z})$ l'estimateur de ces différences de potentiel, on a

$$\inf_{\omega \in \mathcal{E}_n} \mathbb{P}^\omega \left[\bigcap_{k \in \mathbb{L}_n^\gamma} \left\{ |\hat{S}_k^n - S_k^n| < C(\log n)^{-1} \right\} \right] \geq 1 - (\log n)^{-\min(\gamma/2, \gamma-6)}.$$

où $C > 0$ est une constante. Ci-dessous on montre deux simulations, une (à gauche) où l'estimateur est performant sur un large intervalle du fait que la MAMA a largement le temps de visiter une majorité de points (y compris ceux déjà visités) après avoir touché le fond de la $\log n$ -vallée. Une seconde (à droite) où ce n'est plus le cas. En rouge on a tracé S_k^n , en bleu $\hat{S}_x^n - C(\log n)^{-1}$ et en vert $\hat{S}_x^n + C(\log n)^{-1}$.

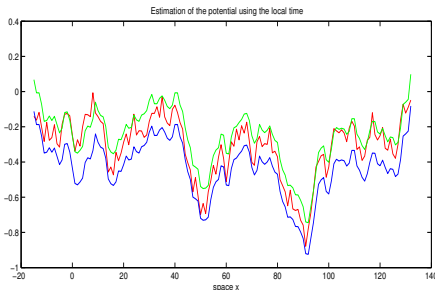


FIGURE 7 – $n = 5.10^5$, $\gamma = 7$, $\mathbb{L}_n^\gamma = [10, 100]$

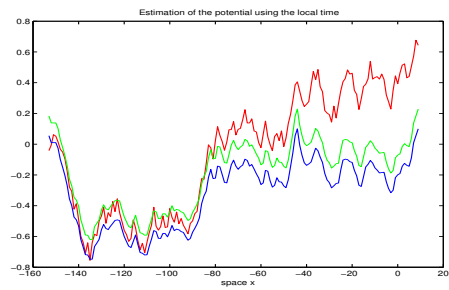


FIGURE 8 – $n = 5.10^5$, $\gamma = 7$, $\mathbb{L}_n^\gamma = [-149, -91]$

▷ Un ancien article (années 60) abondamment cité par les probabilistes ([Che62]) fait état de la modélisation de la réplication de l'ADN par ce qu'on appelle aujourd'hui MAMA. La MAMA introduite dans la partie suivante est également une modélisation plus récente (fin des années 90) de l'ouverture d'une molécule d'ADN, la motivation dans cette approche est le séquençage.

▷

4.2 Dégrafage d'ADN

Les physiciens V. Baldazzi, S. Cocco, E. Marinari et R. Monasson ([BCMM06], [BCMM07], [CM08]), se sont intéressés aux expériences de dégrafage mécanique de molécules d'ADN opérées par les biophysiciens U. Bockelmann, B. Essevaz-Roulet, et F. Heslot [BERH97] et [BERH98]. L'expérience consiste à écarter les deux brins complémentaires d'une molécule d'ADN en appliquant une force f aux extrémités de la molécule (pour plus de précision voir [BCMM06] - voir également le schéma Figure 9). La force f est choisie de manière à ce que la molécule puisse être entièrement dégrafée en fin de processus. Cependant f reste raisonnable afin que la molécule puisse également se reformer partiellement au cours de l'expérience. Ce procédé induit un mouvement de va et vient (ouverture, fermeture) et le nombre de bases ouvertes à chaque instant est mesuré par les biologistes. Le signal obtenu est caricaturé par les physiciens par une chaîne de Markov dont les probabilités de transition sont inconnues.

L'un des objectifs de ce procédé est le séquençage de la molécule, en effet l'énergie de liaison unissant les deux brins dépend fortement des bases dont ils sont composés. Ainsi le lien $A - T$ étant plus faible que le lien $C \equiv G$ pour des raisons de valence (2 liaisons hydrogènes contre 3), le dégrafage sera ralenti ou accéléré suivant la concentration de l'une ou l'autre des paires. Il faut ajouter à cela un effet d'empilement entre bases adjacentes qui contribue à la multiplicité des énergies de liaison entre les bases de la molécule. Autrement dit la force nécessaire pour casser par exemple un lien $C \equiv G$ est différente selon si C est suivi d'un A , d'un T , d'un C ou d'un G . Il s'avère qu'à température ambiante cet effet n'est pas négligeable comme en témoigne le tableau d'énergie de la Figure 10.

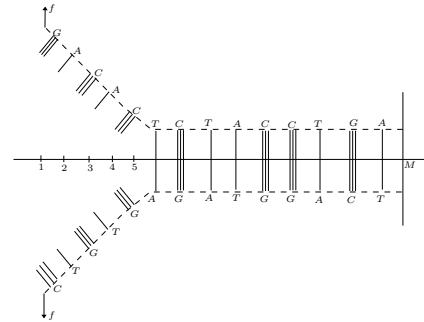


FIGURE 9 – Dégrafage d'ADN

Le modèle

On notera M la longueur de la molécule, c'est à dire le nombre de bases d'un brin de la molécule, et (b_1, b_2, \dots, b_M) la suite de bases ordonnées de ce brin. Ainsi b_i est la i^{th} base à valeurs dans $\{A, T, C, G\}$ par complémentarité on peut en déduire la suite des bases de la molécule. Nous avons considéré deux modèles très similaires l'un noté $(X_n, n \in \mathbb{N})$ à temps discret et le second $(Y(t), t \in \mathbb{R}^+)$ à temps continu. Ces deux processus correspondent au nombre de paires dégrafées à un instant donné.

Nous faisons maintenant le lien entre X (et Y) et b . Commençons par définir l'énergie libre g

| g_0 | A | T | C | G |
|-------|------|------|------|------|
| A | 1.78 | 1.55 | 2.52 | 2.22 |
| T | 1.06 | 1.78 | 2.28 | 2.54 |
| C | 2.54 | 2.22 | 3.14 | 3.85 |
| G | 2.28 | 2.52 | 3.90 | 3.14 |

FIGURE 10 – Energie de Liaison $\times k_B T$, [Zuk00].

associée à la molécule lorsque les x premières bases sont ouvertes :

$$g(x) := \sum_{i=1}^x g_0(b_i, b_{i+1}) - x g_1(f).$$

$g_0(b_i, b_{i+1})$ est l'énergie de liaison de la paire à l'emplacement $i + 1$. Notons que l'effet d'empilement dont nous avons parlé est ici pris en compte car g_0 dépend de b_i et de b_{i+1} . $x g_1(f)$ est l'énergie nécessaire à la force f pour maintenir ouverts les deux brins lorsqu'une nouvelle base est ouverte.

g_1 est connue et fixée alors que $(\sum_{i=1}^x g_0(b_i, b_{i+1}), x \leq M)$ est inconnue et représentera le milieu aléatoire. Une trajectoire typique de g est donnée par la simulation de la Figure 11.

Le nombre de bases ouvertes fluctuent aléatoirement avec une distribution qui va dépendre de la différence entre les énergies libres des bases successives, plus précisément on a :

Le cas (temps) discret (X_n, n) : pour une suite $b := (b_x, 1 \leq x \leq M)$ donnée les probabilités de transition sont données par : pour tout $1 \leq x \leq M - 1$,

$$\mathbb{P}(X_{x+1} = x + 1 | X_x = x, b) = \frac{1}{1 + \exp[\beta(g_0(b_x, b_{x+1}) - g_1(f))]}, \quad (11)$$

où β est une constante proportionnelle à l'inverse de la température. On suppose aussi que la probabilité d'ouvrir la première base vaut 1. Notons que $g_1(f)$ étant croissante avec f , plus celle-ci est grande, plus grande est la probabilité d'ouvrir une nouvelle paire.

Nous reconnaissons donc ici les probabilités de transition d'une marche au plus proche voisin soumis à un potentiel g .

Le cas (temps) continu $(Y(s), s)$: les physiciens prennent aussi en compte le temps d'attente entre deux sauts successifs. Le temps d'attente en une coordonnée $x > 1$ est distribué comme une loi exponentielle de paramètre $e^{-\beta g_0(b_x, b_{x+1})} + e^{-\beta g_1(f)}$ et une loi exponentielle de paramètre $e^{-\beta g_0(b_1, b_2)}$ en $x = 1$. Le processus Y est simplement une chaîne de Markov à temps continu caractérisée par le couple (X, T) où X est la suite de sauts discrets et T la suite de temps d'attente successifs en chaque site. On considèrera ici uniquement le cas où les observations sont données par le processus Y arrêté à l'instant d'atteinte de la fin de la molécule $\tau_M := \inf\{k > 0, X_k = M\} : ((X_k, T_k), k \leq \tau_M)$.

Résultats des physiciens : dans leurs travaux [BCMM06], [BCMM07] et [CM08] étudient d'abord le cas sans effet d'empilement en supposant que $(g_0(b_x), x)$ est une suite de variables aléatoires indépendantes et de même loi. Ils calculent l'Estimateur du Maximum de Vraisemblance \hat{b}_x de b_x . Pour plus de précision ils considèrent R trajectoires indépendantes $(Y^{(l)}, l \leq R)$. Ils montrent que pour $x \leq M$,

$$\mathbb{P}(b_x \neq \hat{b}_x) \leq \exp(-R/R_c(x)),$$

et estiment $R_c(x)$. Pour le cas général (avec effet d'empilement) ils utilisent l'algorithme de Viterbi [Vit67] pour le calcul de l'estimateur. Ils estiment ensuite $\mathbb{P}(b_x \neq \hat{b}_x)$ par des méthodes analytiques et numériques et obtiennent un résultat similaire au cas indépendant.

De discussions avec S. Cocco et R. Monasson sont ressorties les questions suivantes : est-il possible d'obtenir une méthode générale et rigoureuse qui peut être appliquée à tous ces cas ? Comment le choix de la force peut-il être utilisé afin d'améliorer les résultats ? Quelles sont

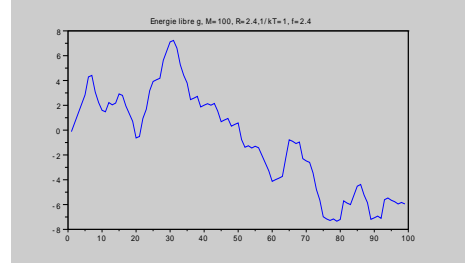


FIGURE 11 – $(g(x), x)$

les différences, du point de vue de l'estimation, entre le modèle à temps discret et le modèle continu ? L'article avec R. Diel [12] a eu pour but de répondre à ces interrogations.

Approche probabiliste : La seule hypothèse que nous ferons ici est le fait que g_0 est injective en ses première et seconde variables : pour tout $a \in \{A, T, C, G\}$

$$g_0(a, \cdot) : \gamma \rightarrow g_0(a, \gamma) \quad \text{et} \quad g_0(\cdot, a) : \gamma \rightarrow g_0(\gamma, a) \text{ sont injectives.} \quad (12)$$

Cette hypothèse est cohérente avec les valeurs expérimentales d'énergie données en Figure 10. On considère donc R trajectoires (dégrafages) indépendantes que l'on note $Y^R := (Y^{(l)}, l \leq R)$. Pour construire l'EMV de la séquence b on a besoin du nombre de pas à droite des processus $Y^{(l)} : \text{pour tout } x \in \llbracket 1, M-1 \rrbracket$

$$L_x^{+, (l)} := \sum_{k=0}^{\tau_M^{(l)}-1} \mathbb{1}_{X_k^{(l)}=x; X_{k+1}^{(l)}=x+1}, \quad L_x^{+, R} := \sum_{l=1}^R L_x^{+, (l)},$$

où $\tau_M^{(l)} := \inf\{k > 0, X_k^{(l)} = M\}$, et les temps d'attente

$$H_x^{(l)} := \sum_{i=1}^{\tau_M^{(l)}} T_i^{(l)} \mathbb{1}_{X_i^{(l)}=x}, \quad H_x^R := \sum_{l=1}^R H_x^{(l)}.$$

• *Estimation base par base :* On commence par une estimation site par site, dans ce cas, l'estimateur du maximum de vraisemblance \hat{b}_x de b_x pour tout $2 \leq x \leq M-1$ est donné par

$$\hat{b}_x = \sum_{\alpha \in \{A, T, C, G\}} \alpha \mathbb{1}_{\{I_x(\alpha, b) = \min_{\bar{\alpha}} I_x(\bar{\alpha}, b)\}},$$

où pour tout $u \in \{A, T, C, G\}$

$$I_x(u, b) := \beta g_0(u, b_{x+1}) L_x^{+, R} + S_x^R e^{-\beta g_0(u, b_{x+1})} + \beta g_0(b_{x-1}, u) L_{x-1}^{+, R} + S_{x-1}^R e^{-\beta g_0(b_{x-1}, u)} - \log \mathbb{P}(b_x = u | b^x),$$

et $b^x := (b_1, b_2, \dots, b_{x-1}, b_{x+1}, \dots, b_M)$. On obtient que $\mathbb{P}.p.s$ pour tout b^x , \hat{b}_x converge lorsque R tend vers l'infini vers b_x . Le résultat de grande déviation auquel s'intéresse les physiciens s'écrit donc :

$$\lim_{R \rightarrow \infty} -\frac{1}{R} \log \mathbb{P}(\hat{b}_x \neq b_x | Y^R, b^x) = \frac{1}{R_c(x)}.$$

Pour l'expression de $R_c(x)$, elle peut être représentée de deux façons.

Représentation quenched de $R_c(x) : \mathbb{P}.p.s. \forall b$

$$\frac{1}{R_c(x)} = \frac{\Delta F^-(b_{x-1})}{\bar{p}_{x-1}} + \frac{\Delta F^+(b_{x+1})}{\bar{p}_x}.$$

où $\Delta F^+(b_{x+1})$ et $\Delta F^-(b_{x-1})$ sont des constantes strictement positives (par l'hypothèse d'injectivité) dépendant uniquement des différences entre deux énergies g_0 et $\bar{p}_z := \mathbb{E}[L_z^{+, (1)} | b^z]$.

Pour la molécule d'ADN on a (via la Figure 10), $\Delta F^-(\gamma) \geq \beta^2 (0.23)^2$, de même pour $\Delta F^+(\gamma)$.

On sait de plus que $(\bar{p}_x)^{-1} \geq \exp(\beta M_x)$, où $M_x := \max_{x+1 \leq l \leq M-1} \left\{ \sum_{k=x+1}^l g_0(b_k, b_{k+1}) - (l-x)g_1(f) \right\}$. Ainsi la convergence est meilleure si il y a une forte concentration de base forte entre x et M associée à une force faible. Pratiquement \bar{p}_x n'est pas connu il est donc intéressant

d'avoir une représentation trajectorielle de la variable $R_c(x)$:

Représentation trajectorielle de $R_c(x)$: \mathbb{P} .p.s. pour tout b et R suffisamment grand

$$\frac{1}{R_c(x)} = \Delta F^-(b_{x-1}) \frac{L_{x-1}^{+,R}}{R} + \Delta F^+(b_{x+1}) \frac{L_x^{+,R}}{R} + \epsilon_x(R),$$

où $\epsilon_x(R) \approx (R \log \log R)^{1/2} L_x^{+, (R)} / R$. Ces résultats confirment ceux des physiciens à savoir que considérer un petit nombre de dégrafage améliore déjà considérablement l'estimation (voir aussi les Simulations Figure 12 où $R = 1$ et Figure 13 avec $R = 10$).

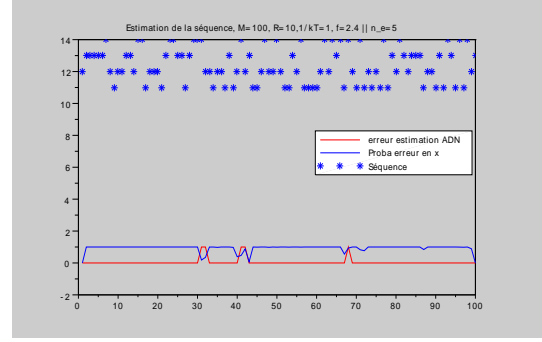
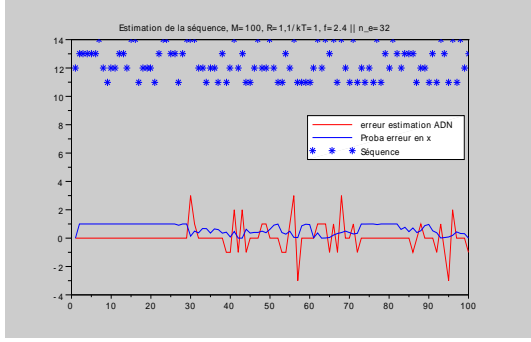


FIGURE 12 – $R=1$, $\hat{b} - b$, $P(\hat{b}_x = b_x | Y^R, b^x)$

FIGURE 13 – $R = 10$, $\hat{b} - b$, $P(\hat{b}_x = b_x | Y^R, b^x)$

On peut reproduire l'analyse en considérant la molécule dans son ensemble :

- *Estimation globale de la molécule* : Pour tout $\alpha \in \{A, T, C, G\}^M$ avec $\alpha_1 = b_1$,

$$\mathbb{P}(b = \alpha | Y^R, b_1) = \frac{e^{-I(\alpha)}}{\sum_{\bar{\alpha}} e^{-I(\bar{\alpha})}},$$

l'estimateur du maximum de vraisemblance global est donc l'élément \tilde{b} de $\{A, T, C, G\}^M$ qui minimise la fonction I :

$$I(\alpha) = -\log \mathbb{P}(b = \alpha) + \sum_{x=1}^{M-1} \beta g_0(\alpha_x, \alpha_{x+1}) L_x^{+,R} + e^{-\beta g_0(\alpha_x, \alpha_{x+1})} S_x^R, \quad \alpha \in \{A, T, C, G\}^M.$$

Théoreme 4.1 (A.-Diel [12]). *Si (12) est satisfaite alors l'estimateur du maximum de vraisemblance global converge presque sûrement lorsque R tend vers l'infini vers la molécule d'ADN b .*

Dans la suite de l'article nous proposons plusieurs façons de moduler la force afin d'améliorer localement la précision du séquençage tout en conservant un temps de dégrafage raisonnable.

▷ Dans cette approche on voit que l'on accorde aucune importance à la distribution du milieu, fort heureusement d'ailleurs car le but est de séquencer n'importe quelle molécule. Si on cherche maintenant à estimer des paramètres intrinsèques à la molécule (les énergies de liaison, la fréquence d'apparition de ces énergies ...) une approche différente est nécessaire. Afin de conserver certaines spécificités de la molécule notamment le fait que les énergies de liaison soient fonction de deux bases successives, on peut conserver le modèle précédent et supposer que le milieu est Markovien. C'est ce qui a motivé le travail qui suit en collaboration avec D. Loukianova et C. Matias, il s'inscrit dans la continuité de récents travaux sur des problèmes d'estimation de la loi de l'environnement à travers des MAMA. ▷

4.3 Estimation en milieu Markovien - Lien entre MAMA et CMC

Si on fait abstraction des articles [AE04], [7] et [11], une majorité de travaux concernant des problèmes d'estimation relatifs aux MAMA sont apparus très récemment. Ils sont effectués par F. Comets, M. Falconnet, A. Gloter, O. Loukianov, D. Loukianova et C. Matias. Ils étudient la consistance et la normalité de l'estimateur du maximum de vraisemblance du paramètre $\theta \in \Theta$ de la distribution de ω_0 (Θ est un compact de $\mathbb{R}^q, q \in \mathbb{N}^*$). Il suppose pour cela que le milieu aléatoire $\omega = (\omega_x)_{x \in \mathbb{Z}} \in \mathcal{S}^{\mathbb{Z}}$ (toujours avec les notations de la Partie 1) est composé de variables i.i.d. Tous les cas sont alors étudiés : le cas transient ballistique ([CFL⁺14], [FLM14]), sous-ballistique ([MAL]) et récurrent ([CFL]).

C'est l'hypothèse d'indépendance que nous allons relaxer ici en supposant que ω est une chaîne de Markov réversible apériodique et Harris récurrente positive. On note donc Q le noyau de transition de ω : pour tout $a \in \mathcal{S}$ et $B \subset \mathcal{S}$

$$Q_\theta(a, B) = \int_B q_\theta(a, b) db,$$

et on désigne par μ_θ la densité de la mesure de probabilité invariante par rapport à la mesure de Lebesgue sur $(0, 1)$ si \mathcal{S} est continue ou par rapport à la mesure de comptage si \mathcal{S} est discrète. \mathbb{P}_θ est la mesure annealed ($\mathbb{P}_\theta(A) := \int dx \mu_\theta(x) \mathbb{P}_\theta(A | \omega_0 = x)$), \mathbb{E}_θ son espérance relative, enfin \mathbb{P}_θ^ω la mesure quenched.

Dans [13] nous traitons le cas où X est transiente ballistique (à droite). L'hypothèse pour obtenir la transience est la même que dans le cas i.i.d. c'est à dire $(\forall \theta \in \Theta) \mathbb{E}_\theta(\log \tilde{\omega}_0) < 0$, tandis que pour la ballisticité

$$\mathcal{R} = (1 + \tilde{\omega}_0 + \tilde{\omega}_0 \tilde{\omega}_1 + \dots)^{-1}$$

doit être de moyenne finie (voir [Ali99]). Comme dans la section précédente nous considérerons les trajectoires finies de $X : \mathbf{X}^n = (X_k)_{k \leq \tau_n}$, (τ_n est toujours le premier temps d'atteinte de la coordonnée $n \in \mathbb{N}$), où par hypothèse $\tau_n < +\infty$, \mathbb{P}_θ -p.s. On suppose que la suite d'observation est générée sous la distribution \mathbb{P}_{θ^*} où θ^* appartient à l'intérieur de Θ . Le but est donc d'estimer θ^* à partir de la suite d'observations \mathbf{X}^n .

• Processus de branchement avec immigrant

L'estimateur de θ^* sera construit à partir du processus des pas à gauche $\mathbf{G} := (G_n^m, G_{n-1}^m, \dots, G_0^m)$ du processus \mathbf{X}^n . Notons que c'est une statistique exhaustive pour le paramètre θ^* comme précisé dans [CFL⁺14].

Un point déterminant ici est le fait suivant (observé dans [KKS75], repris dans [CFL⁺14] et qui reste vrai dans notre cas de milieu Markovien grâce à l'hypothèse de réversibilité) : sous \mathbb{P}_θ , \mathbf{G} a la même distribution que les n premières générations d'un processus de branchement en environnement aléatoire avec un immigrant à chaque génération et loi de reproduction géométrique. Nous noterons donc ce processus $\mathbf{Z} := (Z_0, Z_1, \dots, Z_n)$ tel que $Z_0 = 0$ et $\forall k \geq 0$, $Z_{k+1} = \sum_{i=0}^{Z_k} \xi_{k+1,i}$ où pour tout $k \geq 0$, les variables aléatoires $(\xi_{k,i}, i)$ sont indépendantes de distribution conditionnellement à ω donnée par $\forall m \in \mathbb{N}$, $\mathbb{P}_\theta^\omega(\xi_{k,i} = m) = (1 - \omega_k)^m \omega_k$. L'étude de l'EMV de θ à partir des observations de X se réduit donc à l'étude de l'EMV à partir de celles de \mathbf{Z} .

• Lien avec les Chaînes de Markov Cachées

Notons que lorsque ω est i.i.d. $(Z_k)_{k \geq 0}$ est une chaîne de Markov. Ici ω étant elle même une chaîne de Markov $(Z_k)_{k \geq 0}$ est une chaîne de Markov cachée généralisée. La généralisation venant

du fait que conditionnellement à l'environnement ω (aussi appelé processus latent), le processus observé $(Z_k)_{k \geq 0}$ est Markovien, alors qu'il serait i.i.d. dans le cas de chaînes de Markov cachées classiques⁵. Plus précisément le processus bi-varié $\{(\omega_k, Z_k)\}_{k \geq 0}$ est appelé processus autoregressif avec régime Markovien (voire l'article de Douc-Moulines-Rydén [DMR04]). Il est caractérisé par sa distribution donnée par $\forall(a, b) \in \mathcal{S}^2$ et $\forall(x, y) \in \mathbb{N}^2$,

$$\begin{aligned} \mathbb{P}_\theta(\omega_{k+1} \in db | \omega_k = a) &= q_\theta(a, b)db, \quad \omega_0 \sim \mu_\theta \text{ et} \\ \mathbb{P}_\theta(Z_{k+1} = y | Z_k = x, \omega_{k+1} = a) &= \binom{x+y}{x} a^{x+1} (1-a)^y \text{ et } Z_0 = 0. \end{aligned}$$

Les observations étant données par (Z_0, \dots, Z_n) il est naturel d'estimer la valeur du paramètre θ^* par l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ qui maximise la log-vraisemblance des observations (Z_0, \dots, Z_n) :

$$\hat{\theta}_n \in \underset{\theta \in \Theta}{\text{Argmax}} \log \mathbb{P}_\theta(Z_0, \dots, Z_n).$$

L'étude des propriétés asymptotiques du maximum de vraisemblance pour les Chaînes de Markov Cachées remonte au milieu des années 60 [BP66], pour des résultats récents voir par exemple [DMOvH11]. Pour notre cas de processus auto-régressif avec régime Markovien et espace d'état S borné, les travaux de [DMR04] nous ont servi de base de travail. On montre en particulier que $\{(\omega_k, Z_k)\}_{k \geq 0}$, admet la mesure de probabilité stationnaire π_θ qui est donnée par

$$\forall(a, x) \in S \times \mathbb{N}, \quad \pi_\theta(a, x) = \mu_\theta(a) \mathbb{E}_\theta(\mathcal{R}(1 - \mathcal{R})^x | \omega_0 = a),$$

et aussi que sous l'hypothèse de ballisticité, (ainsi que sous certaines hypothèses techniques sur q_θ et le support de ω_0) la chaîne $\{(\omega_k, Z_k)\}_{k \geq 0}$ est Harris récurrente positive. Nous pouvons ainsi appliquer les résultats de Douc et al. qui montrent la consistance forte de $\hat{\theta}_n$ pour le processus $\{(\omega_k, Z_k)\}_{k \geq 0}$ sous la mesure stationnaire (construite à partir de π_{θ^*}). Ceci implique par couplage, toujours en utilisant les résultats de Douc, la consistance forte de $\hat{\theta}_n$ sous \mathbb{P}_{θ^*} . La normalité est également obtenue en adaptant les hypothèses de [DMR04] à notre cas spécifique, le résultat s'écrit alors

Théoreme 4.2 (A.-Loukianova-Matias [13]). *Sous les hypothèses précitées :*

- 1) *L'estimateur du maximum de vraisemblance $\hat{\theta}_n$ converge \mathbb{P}_{θ^*} presque sûrement vers θ^* .*
- 2) *Sous réserve d'inversibilité de l'information de Fisher $I(\theta^*)$, $n^{-1/2}(\hat{\theta}_n - \theta^*)$ converge en loi vers $\mathcal{N}(0, I^{-1}(\theta^*))$.*

Le second point du théorème nécessite l'inversibilité de l'information de Fisher, un critère d'inversibilité est encore à l'étude au moment où j'écris ces lignes. On donne alors plusieurs exemples typiques que je ne détaillerai pas ici, je reviens par contre sur l'exemple du dégrafage d'ADN.

• **Retour à l'ADN** : On reprend le modèle à temps discret de la Section 4.2. On tient compte du fait que l'énergie de liaison g_0 dépend de deux bases successives en supposant que $(g_0(x) := g_0(b_x, b_{x+1}), x \geq 1)$ est une chaîne de Markov, qui est naturellement apériodique. La réversibilité pour g_0 est vérifiée si on suppose que la distribution de la suite de bases b est identique à la distribution de la suite de bases complémentaires \bar{b} retournée, ceci associé à une propriété de g_0 que l'on peut lire sur le tableau Figure 10, à savoir $g_0(b_1, b_2) = g_0(\bar{b}_2, \bar{b}_1)$ où \bar{b} est la base

5. On connaît le lien fort qui existe entre les MAMA et les marches renforcées (voir par exemple [Pem88], [ES02]), on remarque donc que la communauté statisticienne traite également de MAMA, avec une appellation et des motivations différentes.

complémentaire de b . Ces hypothèses sur g_0 sont transmises aux probabilités de transitions de X via (11). Aussi la condition de ballisticité peut être obtenue naturellement dans la mesure où

$$\tilde{\omega}_x = \exp(\beta(g_0(x) - g_1(f)))$$

et g_1 croît avec f qui peut donc être choisie suffisamment grande de manière à ce que $E_a(\tilde{\omega}_x) < 1$ pour tout $a > 0$ ce qui implique la ballisticité (et la transience).

On remarque aussi que g_0 prend uniquement 10 valeurs (Tableau 10). On est alors dans le cas simple où la chaîne de Markov dont on souhaite estimer les paramètres est à espace discret et on peut appliquer le Théorème précédent au vu des hypothèses. On obtient ainsi les probabilités de passage entre énergies de la molécule. La stationnarité implique aussi que l'on peut estimer μ_θ qui donne les fréquences d'apparition des énergies de liaison. Par l'intermédiaire du Tableau 10, on peut donc aussi retrouver la fréquence d'apparition de certains couples de bases.

4.4 Prolongements

- Dans le cas du milieu Markovien nous sommes, pour le moment, seulement capable d'estimer une partie des paramètres de la distribution de l'environnement, en particulier on ne peut rien dire sur le support de ω_0 . L'article sur lequel nous nous sommes appuyés ne permet pas d'aller plus loin, une approche complémentaire sera donc nécessaire. Aussi nous pouvons nous interroger comme dans le cas du milieu i.i.d. sur les cas non-ballistiques.
- Le dégrafage mécanique de l'ADN et sa modélisation par une MAMA montre à quel point ces objets peuvent s'adapter naturellement à des problèmes de modélisation, c'est aussi un retour aux sources puisque nous gardons en tête le légendaire (pour les probabilistes en tout cas) article de Chernov [Che62]. Pour l'instant uniquement le cas unidimensionnel a été mis en avant. Il se trouve que l'ARN ne se replie pas linéairement mais forme plutôt ce qui ressemblerait localement à un arbre de Galton-Watson [GBH01]. Que peut on dire du lien entre le dégrafage

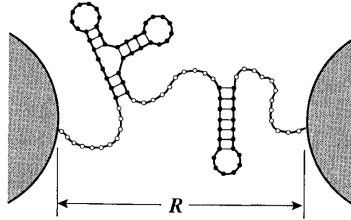


FIGURE 14 – Etirement d'un brin d'ARN – Gerland-Bundschuh-Hwa (2001)

d'un brin d'ARN et des MAMA sur des arbres GW ?

- Dans le cas des diffusions avec pour environnement V un processus de Lévy on pourrait également étudier des problèmes inverses. Par exemple déterminer la hauteur des sauts de V ou encore leur mesure à partir de trajectoires de X .
- Dans [14] je collabore avec des économistes du Laboratoire d'Economie d'Orléans, P. Gazé et M. Menuet, sur un Toy-Model en économie des réseaux. Nous proposons un modèle stochastique (très simple à base de marches aléatoires usuelles) pour un marché-biface. Nous montrons que par rapport au modèle déterministe les profits engendrés par ce modèle de marché sont significativement accrus.

Liste des Travaux

- [1] ANDREOLETTI, P. Alternative proof for the localisation of Sinai's walk. *Journal of Statistical Physics* 118 (2005), 883–933.
- [2] ANDREOLETTI, P. On the concentration of Sinai's walk. *Stoch. Proc. Appl.* 116 (2006), 1377–1408.
- [3] ANDREOLETTI, P. Almost sure estimates for the concentration neighborhood of Sinai's walk. *Stoch. Proc. Appl.* 117 (2007), 1473–1490.
- [4] ANDREOLETTI, P. A limit result for a system of particles in random environment. *J. Stat. Phys.* 131 (2008), 235–246.
- [5] ANDREOLETTI, P. Constants of concentration for a simple recurrent random walk on random environment. *Preprint : arXiv :0803.2006*, 8 pages (2009).
- [6] ANDREOLETTI, P. The local time of a random walk on growing hypercubes. *Preprint : arXiv :0903.2696*, 24 pages (2009).
- [7] ANDREOLETTI, P. On the estimation of the potential of Sinai's rwre. *Braz. J. Probab. Stat.* 25 (2011), 121–235.
- [8] ANDREOLETTI, P., AND DEBS, P. The number of generations entirely visited for recurrent random walks on random environment. *J. Theor. Probab.* 27 (2014), 518–538.
- [9] ANDREOLETTI, P., AND DEBS, P. Spread of visited sites of a random walk along the generations of a branching process. *EJP* 19 (2014), 1–22.
- [10] ANDREOLETTI, P., AND DEVULDER, A. Localization and number of visited valleys for a transient diffusion in random environment. *Preprint : arXiv :1311.6332, soumis*, 52 pages (2014).
- [11] ANDREOLETTI, P., AND DIEL, R. Limit law of the local time for brox's diffusion. *J. Theoretical Probab.* 24 (2011), 634–656.
- [12] ANDREOLETTI, P., AND DIEL, R. Dna unzipping via stopped birth and death processes with random probability transition. *Appl Math Res Express* 2012 (2012), 184–208.
- [13] ANDREOLETTI, P., LOUKIANOVA, D., AND MATIAS, C. Hidden Markov model for parameter estimation of a random walk in a Markov environment. *Preprint : arXiv :1407.4905.pdf, soumis*, 29 pages (2014).
- [14] ANDREOLETTI, P., MENUET, M., AND GAZÉ, P. Can a platform make profit with consumers' mobility ? a two-sided monopoly model with random endogenous side-switching. *Preprint soumis (disponible sur le site du LEO - Orléans)*, 11 pages (2014).

Travaux en cours

- [15] ANDREOLETTI, P., DEVULDER, A., AND VÉCHAMBRE, G. The local time of the slow transient diffusion in a brownian environment.

Références

- [AÖ8] E. Aïdékon. Transient random walks in random environment on a galton–watson tree. *Probab. Theory Related Fields*, **142**(2) : 525–559, 2008.
- [AÖ9] E. Aïdékon. Marches aleatoires en milieu aleatoire et marches branchantes. *Thèse, Université Pierre et Marie Curie, Paris*, 2009.
- [Aï10] E. Aïdékon. Large deviations for transient random walks in random environment on a galton–watson tree. *Ann. Inst. H. Poincaré Probab.*, **46**(2) : 159–189, 2010.
- [Aï11] E. Aïdékon. Speed of the biased random walk on a galton–watson tree. *à paraître Probab. Theory Related Fields*, 2011+.
- [AE04] O. Adelman and N. Enriquez. Random walks in random environment : What a single trajectory tells. *Israel J. Math.*, **142** : 205–220, 2004.
- [AFGH12] G. Ben Arous, G. Fribergh, N. Gantert, and A. Hammond. Biased random walks on a galton–watson tree with leaves. *Ann. Probab.*, **40** : 280–338, 2012.
- [AH12] G. Ben Arous and A. Hammond. Randomly biased walks on subcritical trees. *Comm. Pure Appl. Math.*, **65** : 1481–1527, 2012.
- [AHOZ13] G. Ben Arous, Y. Hu, S. Olla, and O. Zeitouni. Einstein relation for biased random walk on galton–watson trees. *Ann. Inst. H. Poincaré Probab. Statist.*, **49** : 698–721, 2013.
- [Ali99] S. Alili. Asymptotic behaviour for random walks in random environments. *J. Appl. Probab.*, **36**(2) : 334–349, 1999.
- [BCMM06] V. Baldazzi, S. Cocco, E. Marinari, and R. Monasson. Inferring dna sequences from mechanical unzipping : an ideal-case study. *Physical Review Letters E*, **96** : 128102–1–4, 2006.
- [BCMM07] V. Baldazzi, S. Cocco, E. Marinari, and R. Monasson. Inferring dna sequences from mechanical unzipping data : the large-bandwidth case. *Physical Review Letters E*, **75** : 011904–1–33, 2007.
- [BERH97] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot. Molecular stick-slip motion revealed by opening dna with piconewton forces. *Phys. Rev. Let.*, **79** : 4489–4492, 1997.
- [BERH98] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot. Dna strand separation studied by single molecule force measurements. *Phys. Rev. E*, **58** : 2386–2394, 1998.
- [BF08] A. Bovier and A. Faggionato. Spectral analysis of sinai’s walk for small eigenvalues. *Ann. Probab.*, **36** : 198–254, 2008.
- [Big76] J.D. Biggins. The first and last birth problems for a multitype age-dependent branching process. *Adv. Appl. Probab.*, **8** : 446–459, 1976.
- [Big77] J. D. Biggins. Chernoff’s theorem in the branching random walk. *J. Appl. Probability*, **14** : 630–636, 1977.
- [BK97] J.D. Biggins and A.E. Kyprianou. Senata-heyde norming in the branching random walk. *Ann. Probab.*, **25** : 337–360, 1997.

- [BP66] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37 :1554–1563, 1966.
- [Bro86] T. Brox. A one-dimensional diffusion process in a Wiener medium. *Ann. Probab.*, 14(4) : 1206–1218, 1986.
- [Car97] P. Carmona. The mean velocity of a brownian motion in a random lévy potential. *Ann. Probab.*, 25(4) : 1774–1788, 1997.
- [CFL⁺14] F. Comets, M. Falconnet, O. Loukianov, D. Loukianova, and C. Matias. Maximum likelihood estimator consistency for ballistic random walk in a parametric random environment. *Stochastic Processes and Applications*, 124(1) :268–288, 2014.
- [CFL1] F. Comets, M. Falconnet, O. Loukianov, and D. Loukianova. Maximum likelihood estimator consistency for recurrent random walk in a parametric random environment with finite support. *preprint, arxiv 1404.2551*.
- [Che62] A. A. Chernov. Replication of a multicomponent chain, by the "lightning mechanism". *Biophysics*, 12 : 336–341, (1962).
- [Che08] D. Cheliotis. Localization of favorite points for diffusion in random environment. *Stoch. Proc. and Appl.*, 118 :1159–1189, 2008.
- [CM08] S. Cocco and R. Monasson. Reconstructing a random potential from its random walks. *epl*, 81 : 1–6, 2008.
- [Dev06a] A. Devulder. Almost sure asymptotics for a diffusion process in a drifted brownian potential. *Preprint*, 2006.
- [Dev06b] A. Devulder. The maximum of the local time of a diffusion in a drifted brownian potential. *Preprint*, 2006.
- [DGPS07] A. Dembo, N. Gantert, Y. Peres, and Z. Shi. Valleys and the maximal local time for random walk in random environment. *Probability Theory and Related Fields*, 137 : 443–473, 2007.
- [Die11a] R. Diel. Almost sure asymptotics for the local time of a diffusion in brownian environment. *Stoch. Proc. Appl.*, 121 : 2303–2330, 2011.
- [Die11b] R. Diel. Local time of a diffusion in a stable levy environment. *Stochastics*, 83 : 127–152, 2011.
- [DMOvH11] R. Douc, É. Moulines, J. Olsson, and R. van Handel. Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann. Statist.*, 39(1) :474–513, 2011.
- [DMR04] R. Douc, É. Moulines, and T. Rydén. Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.*, 32(5) :2254–2304, 2004.
- [DP91] A. DeMasi and E. Presutti. *Mathematical Methods for Hydrodynamic Limits*. Lectures notes in mathematics, Springer Verlag, 1991.
- [DR13] A. Drewitz and A. F. Ramírez. Selected topics in random walks in random environment. *Preprint, <http://arxiv.org/pdf/1309.2589v1.pdf>*, 2013.

- [Dyn55] E. B. Dynkin. Some limit theorems for sums of independent random quantities with infinite mathematical expectations. *Izv. Akad. Nauk SSSR Ser. Mat.*, 19 : 247–266, 1955.
- [ES02] N. Enriquez and C. Sabot. Edge oriented reinforced random walks and rwre. *C. R. Acad. Sci. Paris, Ser. I*, 335 : 941–946, 2002.
- [ESZ09a] N. Enriquez, C. Sabot, and O. Zindy. Aging and quenched localization for one dimensional random walks in random environment in the sub-ballistic regime. *Bulletin de la Société Mathématique de France*, 137 : 423–452, 2009.
- [ESZ09b] N. Enriquez, C. Sabot, and O. Zindy. Limit laws for transient random walks in random environment on \mathbb{Z} . *Annales de l’Institut Fourier*, 59 : 2469–2508, 2009.
- [ESZ09c] N. Enriquez, C. Sabot, and O. Zindy. A probabilistic representation of constants in kesten’s renewal theorem. *Probability Theory and Related Fields*, 144 : 581–613, 2009.
- [Fag09] A. Faggionato. The alternating marked point process of h -slopes of drifted brownian motion. *Stochastic Processes Appl.*, 119(6) : 1765–1791, 2009.
- [Far11] G. Faraud. A central limit theorem for random walk in a random environment on marked galton-watson trees. *Electronic Journal of Probability*, **16**(6) :174–215, 2011.
- [Fel71] W. Feller. *An Introduction to Probability Theory, Vol. 2*. Wiley, New York, NY, 2nd edition, 1971.
- [FHS12] G. Faraud, Y. Hu, and Z. Shi. Almost sure convergence for stochastically biased random walks on trees. *Probab. Theory Relat. Fields*, **154** :621–660, 2012.
- [FLM14] M. Falconnet, D. Loukianova, and C. Matias. Asymptotic normality and efficiency of the maximum likelihood estimator for the parameter of a ballistic random walk in a random environment. *Mathematical Methods of Statistics*, 23(1) :1–19, 2014.
- [Fri09] A. Fribergh. Marches aléatoires en milieux aléatoires et phénomènes de ralentissement. *Thèse, Université Claude Bernard, Lyon*, 2009.
- [FZ] M. Fang and O. Zeitouni. Consistent minimal displacement of branching random walks. *Electron. Commun. Probab.*, 15.
- [GBH01] U. Gerland, R. Bundschuh, and T. Hwa. Force-induced denaturation of rna. *Biophysical J.*, 2001.
- [Gol84] A. O. Golosov. Localization of random walks in one-dimensional random environments. *Communications in Mathematical Physics*, **92** : 491–506, 1984.
- [GPS10] N. Gantert, Y. Peres, and Z. Shi. The infinite valley for a recurrent random walk in random environment. *Annales de l’Institut Henri Poincaré*, 46 : 525–536, 2010.
- [GS02] N. Gantert and Z. Shi. Many visits to a single site by a transient random walk in random environment. *Stoch. Proc. Appl.*, 99 : 159–176, 2002.
- [HS98] Y. Hu and Z. Shi. The limits of Sinai’s simple random walk in random environment. *Ann. Probab.*, **26**(4) : 1477–1521, 1998.

- [HS07a] Y. Hu and Z. Shi. Slow movement of recurrent random walk in random environment on a regular tree. *Ann. Probab.*, **35** :1978–1997, 2007.
- [HS07b] Y. Hu and Z. Shi. A subdiffusive behavior of recurrent random walk in random environment on a regular tree. *Probab. Theory Related Fields*, **138** :521–549, 2007.
- [HS09] Y. Hu and Z. Shi. Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. *Ann. Probab.*, **37** :742–789, 2009.
- [HS14] Y. Hu and Z. Shi. The potential energy of biased random walks on trees. <http://arxiv.org/pdf/1403.6799.pdf>, 2014+.
- [HSY00] Y. Hu, Z. Shi, and M. Yor. Rates of convergence for one-dimensional diffusion in a drifted brownian potential. *Trans. Amer. Math. Soc.*, 2000.
- [KKS75] H. Kesten, M.V. Kozlov, and F. Spitzer. A limit law for random walk in a random environment. *Comp. Math.*, **30** : 145–168, 1975.
- [KL98] C. Kipnis and C. Landim. *Scaling limits of interacting particle systems*. Springer, 1998.
- [KT97] K. Kawazu and H. Tanaka. A diffusion process in a brownian environment with drift. *J. Math. Soc. Japan*, **49** : 189–211, 1997.
- [LP92] R. Lyons and R. Pemantle. Random walk in a random environment and first-passage percolation on trees. *Annals of Probability*, **20** :125–136, 1992.
- [LPP95] R. Lyons, R. Pemantle, and Y. Peres. Ergodic theory on galton-watson trees : Speed of random walk and dimension of harmonic measure. *Ergod. Th. and Dynam. Sys.*, **15** : 593–619, 1995.
- [LPP96] R. Lyons, R. Pemantle, and Y. Peres. Random walk in a random environment and first-passage percolation on trees. *Probab. Theory Related Fields*, **106** :249–264, 1996.
- [Lyo90] R. Lyons. Random walks and percolation on trees. *Ann. Probab.*, **18** :931–958, 1990.
- [Lyo92] R. Lyons. Random walks, capacity and percolation on trees. *Ann. Probab.*, **20** :2043–2088, 1992.
- [MAL] Falconnet M., Gloter A., and D. Loukianova. Maximum likelihood estimation in the context of a sub-ballistic random walk in a parametric random environment. *preprint, arxiv 1405.2880*.
- [McD95] C. McDiarmid. Minimal position in a branching random walk. *Ann. Appl. Probab.*, **5**(1) : 128–139, 1995.
- [MP02] M.V. Menshikov and D. Petritis. On random walks in random environment on trees and their relationship with multiplicative chaos. *In : Mathematics and Computer Science II*, pages 415–422, 2002.
- [Nev86] J. Neveu. Arbres et processus de galton-watson. *Ann. de l’IHP*, **22** :199–207, 1986.

- [NP89] J. Neveu and J. Pitman. *Renewal property of the extrema and tree property of the excursion of a one-dimensional Brownian motion. Séminaire de Probabilités XXIII*, volume 1372. Springer, Berlin, 1989.
- [Pem88] R. Pemantle. Phase transition in reinforced random walk and rwre on trees. *Ann. of Probab.*, 16 :1229–1241, 1988.
- [Pet75] V.V. Petrov. *Sums of Independent Random Variables*. Springer-Verlag, 1975.
- [PZ08] Y. Peres and O. Zeitouni. A central limit theorem for biased random walks on galton-watson trees. *Probab. Theory Related Fields*, **140** : 595–629, 2008.
- [R88] P. Révész. In random environment the local time can be very big. *Astérisque*, pages 157–158,321–339, 1988.
- [R89] P. Révész. *Random walk in random and non-random environments*. World Scientific, 1989.
- [Sch84] S. Schumacher. Diffusions with random coefficients. *Ph.D. thesis, Univ. of California Los Angeles*, 1984.
- [Sch85] S. Schumacher. Diffusions with random coefficients. *Contemp. Math.*, **41** : 351–356, 1985.
- [Shi98] Z. Shi. A local time curiosity in random environment. *Stoch. Proc. Appl.*, **76**(2) : 231–250, 1998.
- [Shi11] Z. Shi. Random walks and trees. *ESAIM : Proceedings*, **31** : 1–39, (2011).
- [Sin82a] Ya. G. Sinai. The limit behaviour of a one-dimensional random walk in a random medium. *Theory Probab. Appl.*, **27**(2) : 256–268, 1982.
- [Sin82b] Ya. G. Sinai. Lorentz gas and random walk. *Math. Problems in Theor. Physics, Lecture Notes in Physics, Springer Verlag*, 153 :12–14, 1982.
- [Sin07a] A. Singh. Limiting behavior of a diffusion in an asymptotically stable environment. *Ann. Inst. H. Poincaré Probab. Stat.*, **43** : 101–138, 2007.
- [Sin07b] A. Singh. A slow transient diffusion in a drifted stable potential. *Journal of Theoretical Probability*, **20** : 153–166, 2007.
- [Sin08] A. Singh. Rates of convergence of a transient diffusion in a spectrally negative lévy potential. *Ann. Probab.*, **36** : 279–318, 2008.
- [Sol75] F. Solomon. Random walks in random environment. *Ann. Probab.*, **3**(1) : 1–31, 1975.
- [SZ07] Z. Shi and O. Zindy. A weakness in strong localization for sinai’s walk. *Ann. Probab.*, 35 :1118–1140, 2007.
- [Tan94] H. Tanaka. Localization of a diffusion process in a one-dimensional brownian environnement. *Comm. Pure Appl. Math.*, **17** : 755–766, 1994.
- [Tan97] H. Tanaka. Limit theorem for a brownian motion with drift in a white noise environment. *Chaos Solitons Fractals*, **11** : 1807–1816, 1997.

- [Tan04] H. Tanaka. Lévy processes conditioned to stay positive and diffusions in random environment. *Advanced Studies in Pure Math.*, **39**, 2004.
- [Tem69] D. E. Temkin. The theory of diffusionless crystal growth. *Journal of Crystal Growth*, **5**(3) :193–202, 1969.
- [Tem72] D. E. Temkin. One dimensional random walk in two component chain. *Dokl. Akad. Nauk SSSR*, **13**(5) :27–30, 1972.
- [Vit67] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory*, **13**(2) :260–269, 1967.
- [Zei01] O. Zeitouni. Lectures notes on random walks in random environment. *St Flour Summer School*, 2001.
- [Zin08] O. Zindy. Upper limits of sinai’s walk in random scenery. *Stochastic Process. Appl.*, 118 :981–1003, 2008.
- [Zuk00] M. Zuker. Calculating nucleic acid secondary structure. *Curr. Opin. Struct. Biol.*, 10, 2000.

RÉSUMÉ

Dans cette synthèse on trouvera quelques variations autour des marches aléatoires et diffusions browniennes évoluant sur des potentiels aléatoires.

En particulier sont explorés la notion de localisation fluctuante, propre à la dimension un, ainsi que la structure de renouvellement qui apparaît lorsque l'environnement est moins favorable à une localisation forte. Le temps local est aussi étudié en détail, il est le témoin d'une stationnarité, local dans le temps, de ces processus. Il se révélera aussi être l'outil de base utilisé dans les applications.

On se tourne ensuite vers les marches sur potentiels branchants, on s'aperçoit alors que les automatismes de la dimension un doivent être abandonnés pour en comprendre le fonctionnement. On s'attache plus particulièrement à décrire la façon dont l'arbre de Galton-Watson est visité, on découvre alors que la marche peut s'étaler sur l'arbre sans exclure les sites de potentiel élevé.

Finalement on utilise la richesse des marches aléatoires en milieux aléatoires pour décrire les fluctuations du dégrafage mécanique d'une molécule d'ADN. La maîtrise théorique, probabiliste, obtenue dans les parties précédentes permet d'aborder des problèmes d'estimation, qui à leur tour induisent des résultats sur le séquençage et d'autres propriétés intrinsèques à la molécule.