

ÉCOLE NORMALE SUPÉRIEURE DE RENNES

RAPPORT DE STAGE

**La marche aléatoire de l'éléphant**

Alice MORINIÈRE

Stage co-encadré par Hélène Guérin et Lucile Laulin à l'UQAM (Montréal)

Stage effectué en mai-juin 2023

# Table des matières

<b>0 Avant-propos</b>	<b>2</b>
<b>1 Introduction</b>	<b>2</b>
1.1 Rappel sur les marches aléatoires . . . . .	2
1.2 Marche aléatoire de l'éléphant . . . . .	3
<b>2 L'approche martingale</b>	<b>4</b>
2.1 Rapprochement avec les martingales . . . . .	4
2.2 Théorèmes de convergence des martingales . . . . .	7
2.3 Résultats pour le régime diffusif . . . . .	9
2.4 Résultats pour le régime critique . . . . .	13
2.5 Résultats pour le régime super-diffusif . . . . .	14
<b>3 Utilisation des arbres aléatoires récursifs</b>	<b>17</b>
3.1 Définition des arbres aléatoires récursifs et lien avec la marche de l'éléphant . . . . .	18
3.2 Utilisation des arbres pour approfondir le cas super-diffusif . . . . .	20
<b>4 Les urnes de Polya</b>	<b>22</b>
4.1 Lien avec les urnes . . . . .	23
4.2 Utilisation de cette approche . . . . .	24
<b>5 La marche de l'éléphant en dimension supérieure</b>	<b>24</b>
5.1 Définition de la marche en dimension supérieure . . . . .	24
5.2 Les martingales . . . . .	25
5.3 Les arbres aléatoires récursifs . . . . .	27
5.4 Les urnes de Polya . . . . .	28
<b>6 Ouvertures : extensions de la marche aléatoire de l'éléphant</b>	<b>30</b>
6.1 Quelques extensions déjà étudiées . . . . .	30
6.2 De nouvelles pistes... . . . . .	30

# 0 Avant-propos

L'étude suivante portera sur la marche aléatoire de l'éléphant. Le but du stage était surtout d'étudier les différentes modélisations et les résultats obtenus grâce à ces dernières. Ainsi, nous ne rentrerons pas dans les détails des preuves liées à certaines modélisations, qui pour certaines demandes des notions très loin de mes connaissances. La modélisation la plus développée sera donc celle utilisant les martingales. Cependant, même si les preuves n'ont pas été étudiées, la partie sur les arbres aléatoires résulte d'un travail de synthèse et d'adaptation de plusieurs articles plutôt théoriques. Les différentes ouvertures présentes à la fin découlent également d'une réflexion personnelle. Enfin, toutes les simulations présentes par la suite ont été réalisées en Python.

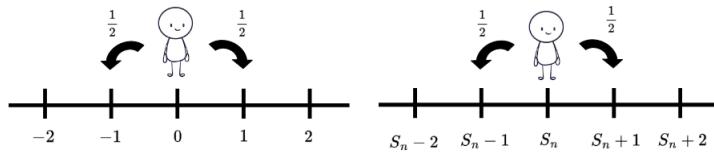
La trame de ce rapport est inspirée de la thèse de Lucile Laulin ([Laulin, 2022c]).

## 1 Introduction

### 1.1 Rappel sur les marches aléatoires

On se place sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Une marche aléatoire est, de façon générale, un modèle mathématique qui se définit par une suite de pas faits aléatoirement, de façon indépendante ou non. La marche aléatoire sur  $\mathbb{Z}$  la plus simple à considérer est la marche aléatoire symétrique. Elle consiste à partir à l'instant initial de 0, et à se déplacer à chaque instant en faisant un pas vers la gauche avec une probabilité  $\frac{1}{2}$  et un pas vers la droite avec cette même probabilité  $\frac{1}{2}$ , chaque pas étant indépendant des précédents.



Ainsi, si pour tout  $n \geq 1$ , on définit  $X_n$  comme étant le  $n$ -ième pas, alors  $(X_n)_{n \geq 1}$  est une famille de variables aléatoires indépendantes et identiquement distribuées suivant comme loi commune une loi Rademacher de paramètre  $\frac{1}{2}$ . Pour tout  $n \geq 0$ , on peut alors définir  $S_n$  la position à l'instant  $n$  comme

$$S_n = \sum_{i=1}^n X_i.$$

Ainsi, la loi des grands nombres et le théorème central limite nous permettent d'obtenir les deux résultats suivants :

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0 \quad \text{et} \quad \frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

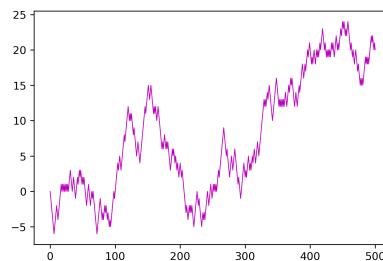
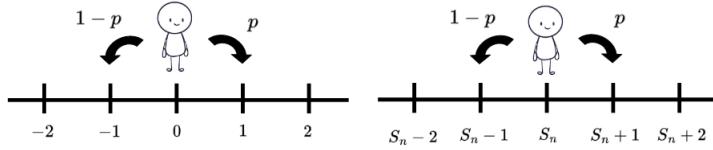


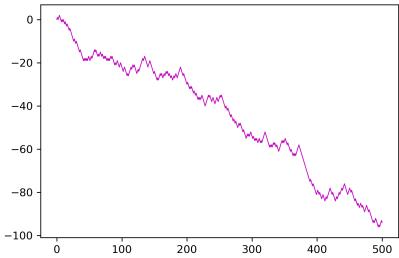
FIGURE 1 – Marche symétrique

Une première amélioration de ce modèle est de considérer que les pas  $(X_n)_{n \geq 0}$  suivent désormais une loi de Rademacher de paramètre  $p$ , pour un certain  $p \in [0, 1]$ .

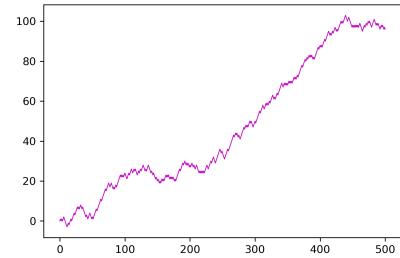


La loi des grands nombres et le théorème central limite nous donnent alors les résultats suivants :

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 2p - 1 \quad \text{et} \quad \frac{S_n - (2p - 1)n}{\sqrt{4p(1-p)n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$



(a) Marche non symétrique pour  $p=0.4$

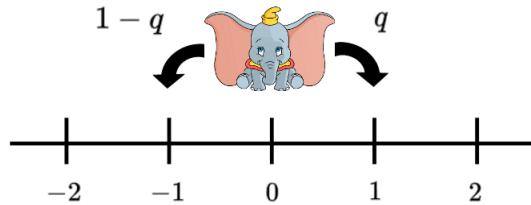


(b) Marche non symétrique pour  $p=0.6$

On pourrait également aller plus loin en regardant cette même marche aléatoire dans  $\mathbb{Z}^d$ .

## 1.2 Marche aléatoire de l'éléphant

La marche aléatoire de l'éléphant est une marche aléatoire avec une mémoire infinie, d'où son nom. Elle se définit de la façon suivante. A l'instant initial, l'éléphant se place en 0. Puis pour son premier pas, il va aller à droite avec une probabilité  $q$ , et à gauche avec une probabilité  $1 - q$ , avec  $q \in [0, 1]$ . Ainsi sa position après le premier pas est donnée par  $S_1 = X_1$ , où  $X_1$  suit une loi de Rademacher de paramètre  $q$ .



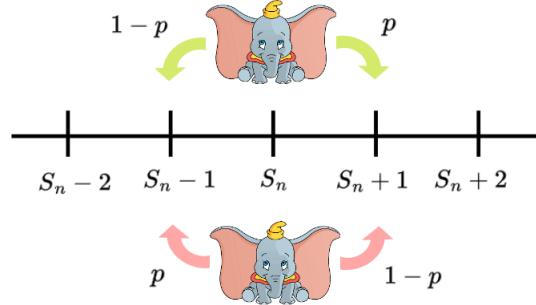
Puis pour toute étape  $n \geq 2$ , l'éléphant va se rappeler uniformément au hasard une étape  $k$  parmi toutes celles qu'il a déjà effectuées, donc  $k$  appartient à  $\llbracket 1, n-1 \rrbracket$ . Ensuite il va faire le même pas qu'à cette étape avec une probabilité  $p$ , et l'inverse avec une probabilité  $1-p$ , où  $p \in [0, 1]$  est le paramètre de mémoire. Ainsi on peut définir

$$X_n = \begin{cases} X_k & \text{avec probabilité } p, \\ -X_k & \text{avec probabilité } 1-p. \end{cases}$$

et alors pour tout  $n \geq 0$ , la position  $S_n$  de l'éléphant à l'étape  $n$  est donnée par

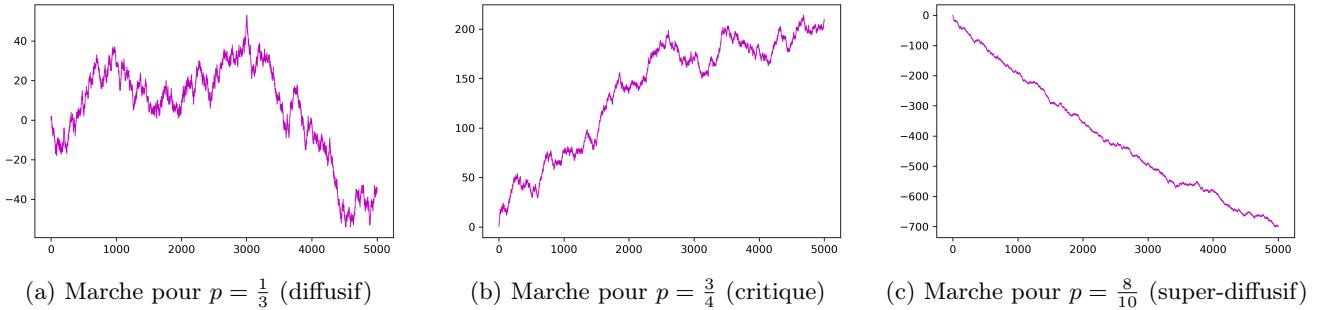
$$S_n = \sum_{i=1}^n X_i.$$

Le cas  $p = \frac{1}{2}$  correspond en fait à la marche aléatoire symétrique définie précédemment.



Dans la figure ci-dessus, l'éléphant se souvient dans le cas vert d'un pas vers la droite, et dans le cas rose d'un pas vers la gauche.

Nous allons voir dans ce qu'il suit que le comportement de cette marche dépend fortement du paramètre de mémoire, et qu'il existe 3 régimes qui sont délimités par la valeur de  $p$  : le régime diffusif, pour  $0 \leq p < \frac{3}{4}$ , le régime critique, pour  $p = \frac{3}{4}$ , et enfin le régime super-diffusif, pour  $\frac{3}{4} < p \leq 1$ . Voici ci-dessous des marches aléatoires de l'éléphant après 5000 pas pour des valeurs de  $p$  dans chaque régime.



Afin d'étudier cette marche, nous allons utiliser plusieurs approches : les martingales, les urnes de Polya et les arbres aléatoires. Toutes ces approches vont nous permettre d'obtenir différents résultats sur le comportement asymptotique de cette marche en fonction du régime. Nous allons pour tout cela nous aider fortement de la thèse de Lucile Laulin [Laulin, 2022c] et de ses différents résultats.

## 2 L'approche martingale

La première approche que nous allons utiliser afin d'obtenir des résultats sur cette marche aléatoire de l'éléphant est l'approche martingale. Elle a été utilisée pour la première fois dans [Coletti et al., 2017] afin d'obtenir une loi des grands nombres commune à tous les régimes, ainsi qu'un théorème central limite pour chacun des régimes. Elle a été ensuite reprise dans [Bercu, 2017] afin d'obtenir des résultats plus fins pour chacun des régimes.

### 2.1 Rapprochement avec les martingales

Tout d'abord, commençons par rappeler ce qu'est une martingale.

**Définition 2.1.1.** Une suite de variables aléatoires  $(M_n)_{n \geq 0}$  est une martingale par rapport à une filtration  $(\mathcal{F}_n)_{n \geq 0}$  si

1. Pour tout  $n \geq 0$ ,  $\mathbb{E}(|M_n|) < \infty$ ,
2. La suite  $(M_n)_{n \geq 0}$  est  $(\mathcal{F}_n)$ -adaptée ,

3. Pour tout  $n \geq 0$ ,  $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n$ .

Nous allons en fait voir que notre marche aléatoire est presque une martingale.

**Proposition 2.1.2.** On pose, pour  $n \geq 1$ ,  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ . On définit  $(M_n)_{n \geq 0}$  par  $M_n := a_n S_n$  pour  $n \geq 0$ , où  $a_0 = a_1 = 1$ , et pour  $n \geq 2$  :

$$a_n := \prod_{k=1}^{n-1} \gamma_k^{-1} = \frac{\Gamma(n)\Gamma(2p)}{\Gamma(n+2p-1)} \sim \frac{\Gamma(2p)}{n^{2p-1}}, \quad \text{où } \gamma_k := \frac{k+2p-1}{k}, \quad (1)$$

en notant  $\Gamma$  la fonction Gamma. Alors  $(M_n)_{n \geq 0}$  est une  $(\mathcal{F}_n)$ -martingale.

*Démonstration.* Pour voir cela, remarquons que l'on peut écrire, pour  $n \geq 1$  :

$$X_{n+1} = \alpha_{n+1} X_{\beta_{n+1}},$$

où  $\alpha_{n+1}$  suit une loi de Rademacher de paramètre  $p$ , et où  $\beta_{n+1}$  suit une loi uniforme sur l'ensemble  $\llbracket 1, n \rrbracket$ ,  $\alpha_{n+1}$  étant indépendante des  $X_1, \dots, X_n$ , et de  $\beta_{n+1}$ .

On a alors

$$\begin{aligned} \mathbb{P}(X_{n+1} = x | X_1, \dots, X_n) &= \sum_{k=1}^n \mathbb{P}_{X_1, \dots, X_n}(\beta_{n+1} = k) \times \mathbb{P}_{X_1, \dots, X_n}(X_{n+1} = x | \beta_{n+1} = k) \quad \text{p.s.} \\ &= \sum_{k=1}^n \frac{1}{n} (\mathbb{1}_{\{X_k=x\}} p + \mathbb{1}_{\{X_k=-x\}} (1-p)) \quad \text{p.s.} \\ &= \frac{1}{2n} \sum_{k=1}^n (1 + (2p-1)X_k x) \quad \text{p.s.} \end{aligned} \quad (2)$$

Et donc

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = (2p-1) \times \frac{1}{n} \sum_{k=1}^n X_k = (2p-1) \frac{S_n}{n} \quad \text{p.s.} \quad (3)$$

Ainsi,

$$\mathbb{E}(S_{n+1} | \mathcal{F}_n) = \mathbb{E}(S_n + X_{n+1} | \mathcal{F}_n) = S_n + (2p-1) \frac{S_n}{n} \quad \text{p.s.}$$

Et cela signifie donc

$$\mathbb{E}(S_{n+1} | \mathcal{F}_n) = \gamma_n S_n \quad \text{p.s., où } \gamma_n = 1 + \frac{2p-1}{n}. \quad (4)$$

Ainsi, si l'on définit  $(M_n)_{n \geq 0}$  comme dans l'énoncé, alors cette suite est adaptée à  $(\mathcal{F}_n)_{n \geq 0}$ ; comme pour  $n \geq 0$ ,  $S_n$  est bornée par  $n$ , on en déduit que  $M_n$  est bornée presque sûrement par  $na_n$  et donc intégrable, et enfin on déduit de ce qui précède que pour  $n \geq 0$ ,

$$\mathbb{E}(M_{n+1} | \mathcal{F}_n) = M_n \quad \text{p.s.}$$

Ainsi,  $(M_n)_{n \geq 0}$  est une  $(\mathcal{F}_n)$ -martingale. □

A terme, ce que nous souhaitons, c'est obtenir des résultats sur la convergence de  $(S_n)_{n \geq 0}$ , nous allons donc pour cela étudier le comportement asymptotique de  $(M_n)_{n \geq 0}$ . Comme il s'agit d'une martingale de carré intégrable (car bornée pour tout  $n$ ), nous allons pour cela nous intéresser à son compensateur  $(\langle M \rangle_n)_{n \geq 0}$ . Rappelons tout d'abord sa définition.

**Définition 2.1.3.** Soit  $(\mathcal{F}_n)_{n \geq 0}$  une filtration, et soit  $(M_n)_{n \geq 0}$  une  $(\mathcal{F}_n)$ -martingale de carré intégrable. On appelle compensateur de  $(M_n)_{n \geq 0}$  la suite  $(\langle M_n \rangle)_{n \geq 0}$  croissante et  $(\mathcal{F}_n)$ -prévisible avec  $\langle M \rangle_0 = 0$ , et pour  $n \geq 1$  donnée par

$$\langle M \rangle_n = \sum_{k=1}^n (\mathbb{E}(M_k^2 | \mathcal{F}_{k-1}) - M_{k-1}^2) = \sum_{k=1}^n \mathbb{E}\left((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}\right).$$

Nous allons maintenant définir une quantité qui nous sera essentielle par la suite, pour l'étude du compensateur, et pour les théorèmes de convergence qui suivront.

**Définition 2.1.4.** On définit pour  $k \geq 1$ ,

$$\varepsilon_k := S_k - \gamma_{k-1} S_{k-1},$$

ainsi pour tout  $n \geq 1$ ,

$$M_n = \sum_{k=1}^n a_k \varepsilon_k,$$

et donc  $\Delta M_n = M_n - M_{n-1} = a_n \varepsilon_n$ .

Nous avons maintenant tous les outils pour calculer le compensateur de notre martingale.

**Proposition 2.1.5.** Pour tout  $n \geq 1$ ,

$$\langle M \rangle_n = \sum_{k=1}^n a_k^2 - (2p-1)^2 \sum_{k=1}^{n-1} a_{k+1}^2 \left(\frac{S_k}{k}\right)^2 \quad \text{p.s.} \quad (5)$$

*Démonstration.* Par définition, le compensateur de  $(M_n)_{n \geq 0}$  est donné par  $\langle M \rangle_0 = 0$ , et pour  $n \geq 1$ ,

$$\langle M \rangle_n = \sum_{k=1}^n \mathbb{E}\left((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}\right) = \sum_{k=1}^n a_k^2 \mathbb{E}(\varepsilon_k^2 | \mathcal{F}_{k-1}).$$

Or d'après (4), pour  $k \geq 2$ ,

$$\mathbb{E}(\varepsilon_k^2 | \mathcal{F}_{k-1}) = \mathbb{E}(S_k^2 | \mathcal{F}_{k-1}) - 2\gamma_{k-1} S_{k-1} \mathbb{E}(S_k | \mathcal{F}_{k-1}) + \gamma_{k-1}^2 S_{k-1}^2 = \mathbb{E}(S_k^2 | \mathcal{F}_{k-1}) - \gamma_{k-1}^2 S_{k-1}^2 \quad \text{p.s.}$$

Et, d'après (3), et comme pour  $k \geq 1$ ,  $\gamma_k = 1 + \frac{2p-1}{n}$ , on a

$$\begin{aligned} \mathbb{E}(S_k^2 | \mathcal{F}_{k-1}) &= S_{k-1}^2 + 2S_{k-1} \mathbb{E}(X_k | \mathcal{F}_{k-1}) + 1 = S_{k-1}^2 + 2S_{k-1}^2 (\gamma_{k-1} - 1) + 1 \quad \text{p.s.} \\ &= 1 + S_{k-1}^2 (2\gamma_{k-1} - 1) \quad \text{p.s.} \end{aligned}$$

Ainsi,

$$\mathbb{E}(\varepsilon_k^2 | \mathcal{F}_{k-1}) = 1 - S_{k-1}^2 (\gamma_{k-1} - 1)^2 = 1 - (2p-1)^2 \left(\frac{S_{k-1}}{k-1}\right)^2 \quad \text{p.s.} \quad (6)$$

Et on obtient alors, pour  $n \geq 1$

$$\langle M \rangle_n = a_1^2 + \sum_{k=2}^n a_k^2 \left(1 - (2p-1)^2 \left(\frac{S_{k-1}}{k-1}\right)^2\right) \quad \text{p.s.}$$

C'est à dire

$$\langle M \rangle_n = \sum_{k=1}^n a_k^2 - (2p-1)^2 \sum_{k=1}^{n-1} a_{k+1}^2 \left(\frac{S_k}{k}\right)^2 \quad \text{p.s.}$$

□

Ainsi le comportement asymptotique de  $(\langle M \rangle_n)_{n \geq 0}$  dépend fortement du comportement en l'infini de la suite  $v_n := \sum_{k=1}^n a_k^2 = \sum_{k=1}^n \left( \frac{\Gamma(k)\Gamma(2p)}{\Gamma(k+2p-1)} \right)^2$ . L'apparition de cette suite permet d'apercevoir la séparation entre les trois régimes. En effet, grâce à des résultats sur le comportement asymptotique de la fonction Gamma, il est possible d'obtenir les trois résultats suivants, que nous admettrons.

**Lemme 2.1.6.** [Admis] Pour  $0 \leq p < \frac{3}{4}$ , c'est à dire pour le régime diffusif, on a

$$\frac{v_n}{n^{3-4p}} \xrightarrow{n \rightarrow \infty} \frac{(\Gamma(2p))^2}{3-4p}. \quad (7)$$

Pour  $p = \frac{3}{4}$ , c'est à dire pour le régime critique, on a

$$\frac{v_n}{\log(n)} \xrightarrow{n \rightarrow \infty} \frac{\pi}{4}. \quad (8)$$

Et enfin pour  $\frac{3}{4} < p \leq 1$ , c'est à dire pour le régime super-diffusif, on a

$$v_n \xrightarrow{n \rightarrow \infty} \sum_{k=1}^{\infty} a_k^2 < +\infty. \quad (9)$$

Ainsi donc, cette approche va nous permettre non seulement de voir la séparation des trois régimes, mais également de prouver des théorèmes de type loi des grands nombres et théorème central limite pour chacun de ces régimes.

Avant de passer à la suite, nous allons observer deux majorations provenant de ce qui précède, et qui nous seront utiles par la suite.

**Lemme 2.1.7.** Pour tout  $p \in [0, 1]$ , on a les majorations presque sûres suivantes :

$$\sup_{n \geq 0} \mathbb{E}(\varepsilon_n^2 | \mathcal{F}_{n-1}) \leq 1 \quad \text{et} \quad \sup_{n \geq 0} \mathbb{E}(\varepsilon_n^4 | \mathcal{F}_{n-1}) \leq \frac{4}{3}.$$

*Démonstration.* L'équation (6) nous permet d'obtenir directement la première inégalité. Ensuite, on a presque sûrement

$$\mathbb{E}(\varepsilon_n^4 | \mathcal{F}_{n-1}) = \mathbb{E}(S_n^4 | \mathcal{F}_{n-1}) - 4\gamma_{n-1}S_{n-1}\mathbb{E}(S_n^3 | \mathcal{F}_{n-1}) + 6(\gamma_{n-1}S_{n-1})^2\mathbb{E}(S_n^2 | \mathcal{F}_{n-1}) - 4(\gamma_{n-1}S_{n-1})^3\mathbb{E}(S_n | \mathcal{F}_{n-1}) + (\gamma_{n-1}S_{n-1})^4$$

Donc grâce au même type de calculs que pour obtenir (6), on obtient

$$\mathbb{E}(\varepsilon_n^4 | \mathcal{F}_{n-1}) = 1 + 2(\gamma_n - 1)^2 S_{n-1}^2 - 3(\gamma_n - 1)^4 S_{n-1}^4 \quad \text{p.s.},$$

c'est à dire

$$\mathbb{E}(\varepsilon_n^4 | \mathcal{F}_{n-1}) = 1 + 2(2p-1)^2 \left( \frac{S_{n-1}}{n-1} \right)^2 - 3(2p-1)^4 \left( \frac{S_{n-1}}{n-1} \right)^4 \quad \text{p.s.}$$

Ainsi, en optimisant le polynôme  $1 + 2x - 3x^2$  pour  $x \in [0, 1]$ , on obtient bien la deuxième majoration souhaitée.  $\square$

## 2.2 Théorèmes de convergence des martingales

Avant de voir les différents résultats que l'on possède pour chacun des régimes, faisons tout d'abord quelques rappels sur les théorèmes de convergence des martingales.

Les deux lois des grands nombres suivantes se retrouvent dans [Duflo, 1997], voir les théorèmes 1.3.15 et 1.3.24.

**Théorème 2.2.1.** [Admis] **Loi des grands nombres pour les martingales** Soit  $(M_n)_{n \geq 0}$  une martingale réelle de carré intégrable et de compensateur  $(\langle M_n \rangle)_{n \geq 0}$ , et on note  $\langle M \rangle_\infty = \lim_n \langle M \rangle_n$ .

1. Si  $\langle M \rangle_\infty < \infty$ , alors  $M_n \xrightarrow[n \rightarrow \infty]{p.s.} M_\infty$ , où  $M_\infty$  est une variable aléatoire finie.
2. Si  $\langle M \rangle_\infty = \infty$ , alors

$$\frac{M_n}{\langle M \rangle_n} \xrightarrow[n \rightarrow \infty]{p.s.} 0,$$

et plus précisément, pour tout  $\gamma > 0$ , on a presque sûrement :

$$\frac{M_n}{\langle M \rangle_n} = o\left(\left(\frac{\log(\langle M \rangle_n)^{1+\gamma}}{\langle M \rangle_n}\right)^{\frac{1}{2}}\right).$$

**Théorème 2.2.2.** [Admis] **Loi des grands nombres pour les séries régressives** Soit  $(\varepsilon_n)_{n \geq 0}$  une suite de variables aléatoires, de carré intégrable et adaptée à la filtration  $(\mathcal{F}_n)_{n \geq 0}$  telle que

$$\forall n \geq 0, \quad \mathbb{E}(\varepsilon_{n+1} | \mathcal{F}_n) = 0 \quad \text{et} \quad \sup_n \mathbb{E}(\varepsilon_{n+1}^2 | \mathcal{F}_n) \leq C \quad p.s.,$$

où  $C$  est une variable aléatoire finie. Soit  $(\phi_n)_{n \geq 0}$  une autre suite de variables aléatoires adaptée à  $\mathcal{F}$ . On pose  $s_n = \sum_{k=0}^n \phi_k^2$  et  $M_n = \sum_{k=1}^n \phi_{k-1} \varepsilon_k$ , martingale adaptée à  $\mathcal{F}$ .

1. Si  $s_\infty < \infty$ , alors  $M_n \xrightarrow[n \rightarrow \infty]{p.s.} M_\infty$ , où  $M_\infty$  est une variable aléatoire finie.
2. Si  $s_\infty = \infty$ , alors

$$\frac{M_n}{s_n} \xrightarrow[n \rightarrow \infty]{p.s.} 0,$$

et plus précisément, pour tout  $\alpha > 0$ , on a

$$M_n^2 = \mathcal{O}((\log(s_n)^{1+\alpha} s_n)) \quad p.s.$$

De plus, si  $C$  est une constante et si pour un  $\beta > 2$ ,  $\sup_n \mathbb{E}(\varepsilon_{n+1}^\beta | \mathcal{F}_n) < \infty$ , on a également

$$M_n^2 = \mathcal{O}((\log(s_n) s_n)) \quad p.s.$$

Le théorème central limite qui suit peut également être trouvé dans [Duflo, 1997], voir corollaire 2.1.10.

**Théorème 2.2.3.** [Admis] **Théorème Central Limite pour les martingales** Soit  $(M_n)_{n \geq 0}$  une martingale de carré intégrable adaptée à  $\mathcal{F}$  de compensateur  $(\langle M_n \rangle)_{n \geq 0}$ . On pose, pour  $n \geq 1$   $\Delta M_n = M_n - M_{n-1}$ . Soit  $(s_n)_{n \geq 0}$  une suite réelle déterministe, qui est croissante vers l'infini, et telle que les hypothèses suivantes soient vérifiées :

1.  $\frac{\langle M_n \rangle}{s_n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \lambda$ ,
2. La condition de Lindeberg est vérifiée, i.e. pour tout  $\varepsilon > 0$ ,

$$\frac{1}{s_n} \sum_{k=1}^n \mathbb{E}(|\Delta M_n|^2 \mathbf{1}_{\{|\Delta M_n| \geq \varepsilon \sqrt{s_n}\}} | \mathcal{F}_{k-1}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Alors

$$\frac{M_n}{s_n} \xrightarrow[n \rightarrow \infty]{p.s.} 0 \quad \text{et} \quad \frac{M_n}{\sqrt{s_n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \lambda).$$

Le théorème suivant est un autre théorème central limite pour les martingales, qui est une version simplifiée de celui que l'on peut fois retrouver dans [Heyde, 1977], voir corollaire 1.b.

**Théorème 2.2.4.** *[Admis] Soit  $(M_n)_{n \geq 0}$  une martingale de moyenne nulle et de carré intégrable. Soit  $d_k = M_k - M_{k-1}$ , pour  $k \geq 1$ , avec  $M_0 = 0$ . Si  $\sum_{k=1}^{\infty} \mathbb{E}(d_k^2) < \infty$ , alors on pose  $s_n^2 = \sum_{k=n}^{\infty} \mathbb{E}(d_k^2)$  pour  $n \geq 1$ , et on a*

- La limite  $M_{\infty} = \sum_{k=1}^{\infty} d_k$  existe presque sûrement et  $M_n \xrightarrow[n \rightarrow \infty]{\mathbb{L}^2} M_{\infty}$ .
- Et si on a de plus

$$1. \frac{1}{s_n^2} \sum_{k=n}^{\infty} \mathbb{E}(d_k^2 | \mathcal{F}_{k-1}) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1,$$

$$2. \forall \varepsilon > 0, \frac{1}{s_n^2} \sum_{k=n}^{\infty} \mathbb{E}(d_k^2 \mathbf{1}_{\{|d_k| > \varepsilon s_n\}}) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0,$$

alors

$$\frac{1}{s_{n+1}} \sum_{k=n+1}^{\infty} d_k \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

c'est à dire

$$\frac{M_{\infty} - M_n}{s_{n+1}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Maintenant que l'on a rappelé ces théorèmes, nous allons donc essayer de montrer les différents théorèmes pour chaque régime résumés dans le tableau suivant, où  $L$  est une variable aléatoire non-dégénérée :

	diffusif ( $0 \leq p < \frac{3}{4}$ )	critique ( $p = \frac{3}{4}$ )	super-diffusif ( $\frac{3}{4} < p \leq 1$ )
<b>LGN</b>	$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0$	$\frac{S_n}{\sqrt{n} \log(n)} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0$	$\frac{S_n}{n^{2p-1}} \xrightarrow[n \rightarrow \infty]{\text{p.s., } \mathbb{L}^4} L$
<b>TCL</b>	$\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{3-4p}\right)$	$\frac{S_n}{\sqrt{n} \log(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$	$\frac{S_n - n^{2p-1}L}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4p-3}\right)$

### 2.3 Résultats pour le régime diffusif

Le régime diffusif correspond au cas  $0 \leq p < \frac{3}{4}$ . On peut dans ce cas démontrer la loi des grands nombres suivante, que l'on peut retrouver dans la référence [Bercu, 2017].

**Théorème 2.3.1.** *Pour tout  $p \in [0, \frac{3}{4}[$ , on a*

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

*Démonstration.* Pour prouver ce théorème, nous allons nous servir du théorème 2.2.2. On a par définition  $\mathbb{E}(\varepsilon_{n+1} | \mathcal{F}_n) = 0$  presque sûrement, et le lemme 2.1.7 nous donne la seconde condition sur  $(\varepsilon_n)_{n \geq 0}$ . La suite  $(v_n)_{n \geq 0}$  correspond à la suite  $(s_n)_{n \geq 0}$  du théorème 2.2.2, or on sait grâce à (8) que  $(v_n)_{n \geq 0}$  tend vers l'infini. On obtient donc grâce à ce même théorème :

$$\frac{M_n}{v_n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

Mais cela ne suffit pas à obtenir le résultat souhaité. Cependant le théorème nous donne plus fort que cela, car  $\sup_n \mathbb{E}(\varepsilon_{n+1}^2 | \mathcal{F}_n)$  est bien majorée presque sûrement par une constante, et on sait toujours d'après le lemme 2.1.7, que  $\sup_n \mathbb{E}(\varepsilon_{n+1}^4 | \mathcal{F}_n) < \infty$  presque sûrement :

$$\frac{M_n^2}{v_n} = \mathcal{O}(\log(v_n)) \quad \text{p.s.}$$

et donc d'après (7)

$$\frac{M_n^2}{n^{3-4p}} = \mathcal{O}(\log(n)) \quad \text{p.s.}$$

Et comme pour  $n \geq 0$ ,  $M_n = a_n S_n$ , et  $a_n \sim \frac{\Gamma(2p)}{n^{2p-1}}$ , on obtient

$$\frac{S_n^2}{n} = \mathcal{O}(\log(n)) \quad \text{p.s.}, \quad (10)$$

et donc

$$\left( \frac{S_n}{n} \right)^2 \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0,$$

d'où le résultat.  $\square$

On peut illustrer ce résultat par le grapique d'une simulation.

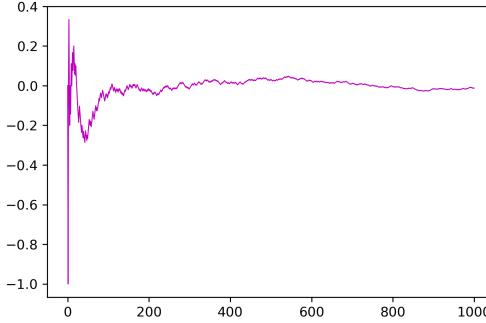


FIGURE 4 – Evolution de  $\frac{S_n}{n}$  en fonction de  $n$ , pour  $p = \frac{1}{3}$

En fait, si l'on compare à l'aide du tableau récapitulatif de la partie précédente, ce résultat laisse penser que ce régime seulement diffusif se diffuse plus que le régime critique ou le régime super-diffusif. Mais cela est dû au fait que ce résultat n'est pas assez fin. Il est effectivement possible de raffiner un peu ce résultat afin de mieux voir le caractère moins diffusif de ce régime.

**Proposition 2.3.2.** Pour tout  $p \in [0, \frac{3}{4}[$ , on a, pour tout  $\varepsilon > 0$ ,

$$\frac{S_n}{\sqrt{n}(\log(n))^{\frac{1}{2}+\varepsilon}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

*Démonstration.* Soit  $\varepsilon > 0$ . On reprend l'équation (10), et on divise par  $(\log(n))^{1+2\varepsilon}$ , on obtient alors

$$\frac{S_n^2}{n(\log(n))^{1+2\varepsilon}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

Et ainsi, en prenant la racine, on obtient bien

$$\frac{S_n}{\sqrt{n}(\log(n))^{\frac{1}{2}+\varepsilon}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

$\square$

De nouveau, on peut observer cette loi des grands nombres grâce à une simulation.

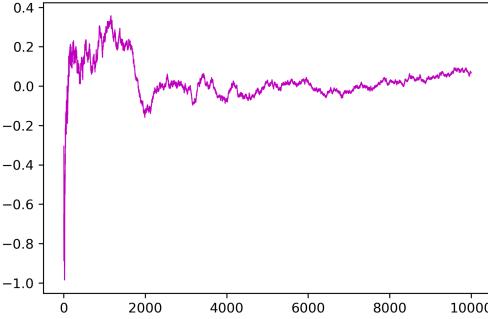


FIGURE 5 – Evolution de  $\frac{S_n}{\sqrt{n}(\log(n))^{\frac{1}{2}+\varepsilon}}$  en fonction de  $n$ , pour  $p = \frac{1}{3}$  et  $\varepsilon = 0.2$

En plus de ces lois des grands nombres, on dispose d'un théorème central limite, que l'on peut de nouveau retrouver dans [Bercu, 2017].

**Théorème 2.3.3.** *Pour tout  $p \in [0, \frac{3}{4}[$ , on a*

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{3-4p}\right).$$

On rappelle le lemme de Kronecker dont nous allons nous servir par la suite.

**Lemme 2.3.4.** *Soient  $(x_n)_{n \geq 1}$  et  $(b_n)_{n \geq 1}$  des suites de réels telles que pour  $n \geq 1$ ,  $b_n > 0$ ,  $(b_n)_{n \geq 1}$  croit vers l'infini et  $\sum_k \frac{x_k}{b_k}$  converge. Alors*

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{k=1}^n x_k = 0.$$

Passons maintenant à la preuve du théorème.

*Démonstration.* On va pour cela utiliser le théorème 2.2.3. Tout d'abord, l'équation (5) nous permet de déduire

$$\frac{\langle M \rangle_n}{v_n} = 1 - (2p-1)^2 \times \frac{1}{\sum_{k=1}^n a_k^2} \times \sum_{k=1}^{n-1} a_{k+1}^2 \left( \frac{S_k}{k} \right)^2 \quad \text{p.s.,} \quad (11)$$

et d'après (7) et le théorème 2.3.1,

$$\frac{\langle M \rangle_n}{v_n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1, \quad (12)$$

ce qui nous permet de vérifier la première condition. Il reste maintenant à vérifier la condition de Linderberg. Soit  $\varepsilon > 0$ . On doit vérifier :

$$\frac{1}{v_n} \sum_{k=1}^n \mathbb{E} (|\Delta M_k|^2 \mathbb{1}_{\{|\Delta M_k| \geq \varepsilon \sqrt{v_n}\}} | \mathcal{F}_{k-1}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Or

$$\begin{aligned}
\frac{1}{v_n} \sum_{k=1}^n \mathbb{E}(|\Delta M_k|^2 \mathbb{1}_{\{|\Delta M_k| \geq \varepsilon \sqrt{v_n}\}} | \mathcal{F}_{k-1}) &= \frac{1}{v_n} \sum_{k=1}^n \mathbb{E}\left(|\Delta M_k|^4 \times \frac{1}{|\Delta M_k|^2} \mathbb{1}_{\{|\Delta M_k| \geq \varepsilon \sqrt{v_n}\}} | \mathcal{F}_{k-1}\right) \quad \text{p.s.} \\
&\leq \frac{1}{\varepsilon^2 v_n^2} \sum_{k=1}^n \mathbb{E}(|\Delta M_k|^4 | \mathcal{F}_{k-1}) \quad \text{p.s.} \\
&\leq \sup_{1 \leq k \leq n} \mathbb{E}(\varepsilon_k^4 | \mathcal{F}_{k-1}) \times \frac{1}{\varepsilon^2 v_n^2} \sum_{k=1}^n a_k^4 \quad \text{p.s.}
\end{aligned}$$

Et donc d'après le lemme 2.1.7, on obtient

$$\frac{1}{v_n} \sum_{k=1}^n \mathbb{E}(|\Delta M_k|^2 \mathbb{1}_{\{|\Delta M_k| \geq \varepsilon \sqrt{v_n}\}} | \mathcal{F}_{k-1}) \leq \frac{4}{3\varepsilon^2 v_n^2} \sum_{k=1}^n a_k^4 \quad \text{p.s.}$$

Mais  $\sum_{k=1}^{\infty} \frac{a_k^4}{v_k^2} < \infty$  car d'après (1) et (8),  $\frac{a_k^4}{v_k^2} \sim \frac{(3-4p)^2}{k^2}$ , on a donc affaire avec une série de Riemann, qui converge.

Ainsi  $\sum_k \frac{a_k^4}{v_k^2}$  converge, et  $(v_k^2)_{k \geq 1}$  croît vers l'infini, donc d'après le lemme de Kronecker,

$$\lim_{n \rightarrow \infty} \frac{1}{v_n^2} \sum_{k=1}^n a_k^4 = 0,$$

ce qui achève la vérification de la condition de Lindeberg. Le théorème 2.2.3 nous donne donc

$$\frac{1}{\sqrt{v_n}} M_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Mais  $M_n = a_n S_n$ , et d'après (1) et (8),  $\frac{a_n}{\sqrt{v_n}} \sim \frac{\sqrt{3-4p}}{\sqrt{n}}$ , et donc

$$\sqrt{3-4p} \times \frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

d'où le résultat. □

Voici le résultat d'une autre simulation illustrant le théorème central limite que l'on vient d'obtenir.

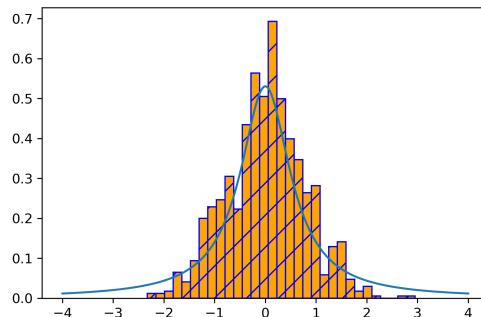


FIGURE 6 – Comparaison entre l'histogramme d'un échantillon de 1000 valeurs de  $\frac{S_n}{\sqrt{n}}$  pour  $n = 500$ , et  $p = \frac{1}{3}$  et la représentation de la densité de la loi  $\mathcal{N}\left(0, \frac{1}{3-4p}\right)$

## 2.4 Résultats pour le régime critique

Passons maintenant au régime critique, qui correspond au cas de  $p = \frac{3}{4}$ . De même que pour le régime diffusif, on dispose d'une loi des grands nombres et d'un théorème central limite, que l'on peut de nouveau retrouver dans [Bercu, 2017].

**Théorème 2.4.1.** *On suppose  $p = \frac{3}{4}$ , alors*

$$\frac{S_n}{\sqrt{n} \log(n)} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

*Démonstration.* On va de nouveau se servir du théorème 2.2.2. Les conditions sur  $(\varepsilon_n)_{n \geq 0}$  sont toujours vérifiées, et (8) nous assure  $v_n \sim \frac{\pi}{4} \log(n)$  et donc  $\lim_{n \rightarrow \infty} v_n = \infty$ , ainsi le théorème nous donne

$$\frac{M_n^2}{\log(n)} = \mathcal{O}(\log(\log(n))) \quad \text{p.s.,}$$

et comme  $M_n = a_n S_n$ , avec  $a_n^2 \sim \frac{\pi}{n}$ , on a

$$\frac{S_n^2}{n \log(n)} = \mathcal{O}(\log(\log(n))) \quad \text{p.s..}$$

Et on en déduit

$$\frac{S_n^2}{n \log(n)^2} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0,$$

d'où le résultat.  $\square$

Illustrons ce résultat à l'aide d'une simulation.

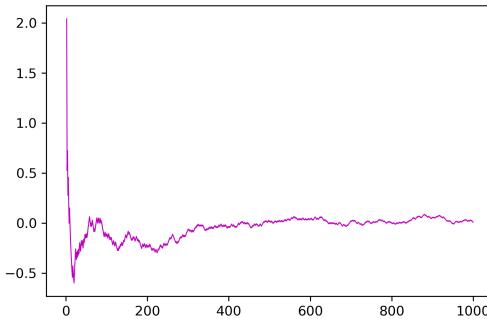


FIGURE 7 – Evolution de  $\frac{S_n}{\sqrt{n} \log(n)}$  en fonction de  $n$ , pour  $p = \frac{3}{4}$

Passons maintenant au théorème central limite.

**Théorème 2.4.2.** *On suppose  $p = \frac{3}{4}$ , alors*

$$\frac{S_n}{\sqrt{n} \log(n)} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

*Démonstration.* On va de nouveau se servir du théorème 2.2.3. L'égalité (11) est toujours valable, et on peut déduire de la preuve qui précède que pour ce régime on a toujours  $\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0$ . Ainsi, grâce à cela et à (8), on peut déduire

$$\frac{\langle M \rangle_n}{v_n} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1,$$

ce qui vérifie la première hypothèse. On a vu précédemment que la condition de Lindeberg était vérifiée dès que  $\lim_{n \rightarrow \infty} \frac{1}{v_n^2} \sum_{k=1}^n a_k^4 = 0$ . Mais  $\sum_k \frac{a_k^4}{v_k^2} < \infty$ , car d'après (1) et (8),  $\frac{a_k^2}{v_k} \sim \frac{1}{n \log(n)}$ , et la série  $\sum_n \frac{1}{n^2 \log(n)^2}$  converge, donc  $\sum_k \frac{a_k^4}{v_k^2} < \infty$  et  $(v_k^2)_{k \geq 1}$  croit vers l'infini, donc d'après le lemme de Kronecker,  $\lim_{n \rightarrow \infty} \frac{1}{v_n^2} \sum_{k=1}^n a_k^4 = 0$ . Ainsi la condition de Linderberg est vérifiée, et donc on peut appliquer le théorème 2.2.3. On a donc

$$\frac{M_n}{\sqrt{v_n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

et comme  $M_n = a_n S_n$ , et  $\frac{a_n}{\sqrt{v_n}} \sim \frac{1}{\sqrt{n \log(n)}}$ , on obtient directement

$$\frac{S_n}{\sqrt{n \log(n)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

□

Voici de nouveau une simulation illustrant le théorème que l'on vient de prouver.

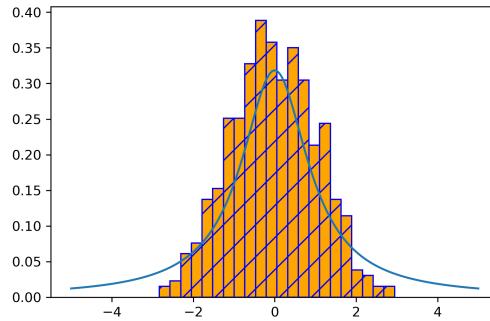


FIGURE 8 – Comparaison entre l'histogramme d'un échantillon de 1000 valeurs de  $\frac{S_n}{\sqrt{n \log(n)}}$  pour  $n = 500$ , et  $p = \frac{3}{4}$  et la représentation de la densité de la loi  $\mathcal{N}(0, 1)$

## 2.5 Résultats pour le régime super-diffusif

Le troisième et dernier régime est le régime super-diffusif qui correspond au cas  $\frac{3}{4} < p \leq 1$ . Comme dans les cas précédents, l'approche martingale nous permet d'obtenir une loi des grands nombres, que l'on peut retrouver dans [Bercu, 2017].

**Théorème 2.5.1.** *Pour tout  $p \in ]\frac{3}{4}, 1]$ , on a*

$$\frac{S_n}{n^{2p-1}} \xrightarrow[n \rightarrow \infty]{p.s.} L, \quad (13)$$

où  $L$  est une variable aléatoire non dégénérée, et cette convergence est également vraie dans  $\mathbb{L}^4$ , c'est à dire

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \left| \frac{S_n}{n^{2p-1}} - L \right|^4 \right) = 0.$$

*Démonstration.* On sait d'après (9) que pour ce régime,  $(v_n)_{n \geq 0}$  possède une limite finie, et on voit dans l'équation (5) que pour tout  $n \geq 0$ ,  $\langle M \rangle_n \leq v_n$ , et donc  $\langle M \rangle_\infty < \infty$ . Ainsi, on applique le théorème 2.2.1, et on obtient

$$M_n \xrightarrow[n \rightarrow \infty]{p.s.} M, \quad \text{avec} \quad M := \sum_{k=1}^{\infty} a_k \varepsilon_k.$$

Et comme  $M_n = a_n S_n$ , et  $a_n \sim \frac{\Gamma(2p)}{n^{2p-1}}$  d'après (1), on obtient directement

$$\frac{S_n}{n^{2p-1}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} L, \quad \text{avec} \quad L := \frac{1}{\Gamma(2p)} \sum_{k=1}^{\infty} a_k \varepsilon_k.$$

Pour prouver cette convergence dans  $\mathbb{L}^4$ , il suffit de montrer que  $(M_n)_{n \geq 0}$  est bornée dans  $\mathbb{L}^4$ , et le théorème de convergence dominée justifie le reste. Montrer tout d'abord que  $(M_n)_{n \geq 0}$  est bornée dans  $\mathbb{L}^2$ , ce qui nous sera utile par la suite. En effet on a  $M_{n+1} = a_{n+1} \varepsilon_{n+1} + M_n$ , et donc

$$\mathbb{E}(M_{n+1}^2 | \mathcal{F}_n) = a_{n+1}^2 \mathbb{E}(\varepsilon_{n+1}^2 | \mathcal{F}_n) + M_n^2 \leq a_{n+1}^2 + M_n^2 \quad \text{p.s.},$$

et donc en passant à l'espérance, on a

$$\mathbb{E}(M_{n+1}^2) \leq a_{n+1}^2 + \mathbb{E}(M_n^2).$$

Et ainsi, par récurrence on obtient

$$\mathbb{E}(M_n^2) \leq \sum_{k=1}^n a_k^2, \tag{14}$$

et ainsi, d'après (9), cela implique

$$\sup_n \mathbb{E}(M_n^2) \leq \sum_{k=1}^{\infty} a_k^2 < \infty.$$

On a donc montré que  $(M_n)_{n \geq 0}$  est bornée dans  $\mathbb{L}^2$ . Montrons maintenant que  $(M_n)_{n \geq 0}$  est bornée dans  $\mathbb{L}^4$ .

On a

$$\mathbb{E}(M_{n+1}^4 | \mathcal{F}_n) = M_n^4 + 4M_n^3 a_{n+1} \mathbb{E}(\varepsilon_{n+1} | \mathcal{F}_n) + 6M_n^2 a_{n+1}^2 \mathbb{E}(\varepsilon_{n+1}^2 | \mathcal{F}_n) + 4M_n a_{n+1}^3 \mathbb{E}(\varepsilon_{n+1}^3 | \mathcal{F}_n) + a_{n+1}^4 \mathbb{E}(\varepsilon_{n+1}^4 | \mathcal{F}_n) \quad \text{p.s.} \tag{15}$$

Or on sait que  $\mathbb{E}(\varepsilon_{n+1} | \mathcal{F}_n) = 0$ , et on a déjà vu dans le lemme 2.1.7

$$\sup_{n \geq 0} \mathbb{E}(\varepsilon_n^2 | \mathcal{F}_{n-1}) \leq 1 \quad \text{et} \quad \sup_{n \geq 0} \mathbb{E}(\varepsilon_n^4 | \mathcal{F}_{n-1}) \leq \frac{4}{3} \quad \text{p.s.}$$

Grâce au même type de calculs que pour ce lemme, on obtient

$$\begin{aligned} \mathbb{E}(\varepsilon_{n+1}^3 | \mathcal{F}_n) &= 2S_n^3 (\gamma_n - 1)^3 - 2S_n (\gamma_n - 1) \quad \text{p.s.} \\ &= 2(2p-1) \frac{S_n}{n} \left( (2p-1)^2 \left( \frac{S_n}{n} \right)^2 - 1 \right) \quad \text{p.s.} \end{aligned}$$

Ainsi

$$M_n \mathbb{E}(\varepsilon_{n+1}^3 | \mathcal{F}_n) = 2(2p-1) a_n \frac{S_n^2}{n} \left( (2p-1)^2 \left( \frac{S_n}{n} \right)^2 - 1 \right) \leq 0 \quad \text{p.s.}$$

Ainsi, en reprenant (15), et en utilisant tout cela, on déduit

$$\mathbb{E}(M_{n+1}^4 | \mathcal{F}_n) \leq \frac{4}{3} a_{n+1}^4 + 6a_{n+1}^2 M_n^2 + M_n^4 \quad \text{p.s.}$$

On passe à l'espérance de cette inéquation et on parvient à l'inégalité suivante

$$\mathbb{E}(M_{n+1}^4) \leq \frac{4}{3} a_{n+1}^4 + 6a_{n+1}^2 \mathbb{E}(M_n^2) + \mathbb{E}(M_n^4).$$

Ainsi, par récurrence, et d'après (14),

$$\begin{aligned}\mathbb{E}(M_n^4) &\leq \frac{4}{3} \sum_{k=1}^n a_k^4 + 6 \sum_{k=1}^n a_k^2 \mathbb{E}(M_{k-1}^2) \\ &\leq \frac{4}{3} \sum_{k=1}^n a_k^2 + 6 \left( \sum_{k=1}^n a_k^2 \right)^2 \\ &\leq 6 \sum_{k=1}^n a_k^2 \left( 1 + \sum_{k=1}^n a_k^2 \right),\end{aligned}$$

et ainsi, toujours grâce à 9, on en déduit que  $(M_n)_{n \geq 0}$  est bornée dans  $\mathbb{L}^4$ , et donc que la convergence a également lieu dans  $\mathbb{L}^4$ .  $\square$

Comme pour les autres régimes, on possède, en plus de cette loi des grands nombres, un théorème central limite pour le régime super-diffusif, que l'on peut cette fois retrouver dans [Kubota and Takei, 2019].

**Théorème 2.5.2.** *Pour tout  $p \in ]\frac{3}{4}, 1]$ , on a*

$$\frac{S_n - n^{2p-1}L}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left( 0, \frac{1}{4p-3} \right).$$

*Démonstration.* Pour prouver cela nous allons utiliser le théorème 2.2.4. Posons pour  $n \geq 0$ ,  $\tilde{M}_n = a_n(S_n - \mathbb{E}(S_n))$ . Alors  $(\tilde{M}_n)_{n \geq 0}$  est toujours une martingale de carré intégrable, et qui est de moyenne nulle. On pose donc  $d_k = \tilde{M}_k - \tilde{M}_{k-1}$  pour  $k \geq 1$ , et on veut  $\sum_{k=1}^{\infty} \mathbb{E}(d_k^2) < \infty$ . Or d'après (4), on a pour  $k \geq 1$

$$\begin{aligned}d_k &= a_k S_k - a_k \gamma_{k-1} \mathbb{E}(S_{k-1}) - a_{k-1} S_{k-1} + a_{k-1} \mathbb{E}(S_{k-1}) \quad \text{p.s.} \\ &= a_k (S_k - \mathbb{E}(S_k | \mathcal{F}_{k-1})) = a_k (X_k - \mathbb{E}(X_k | \mathcal{F}_{k-1})) \quad \text{p.s.}\end{aligned}$$

On remarque que pour  $k \geq 1$ ,  $|d_k| \leq 2a_k$ , et donc  $\mathbb{E}(d_k^2) \leq 4a_k^2$ . Et on déduit de (9) que,  $\sum_{k=1}^{\infty} \mathbb{E}(d_k^2) \leq 4 \sum_{k=1}^{\infty} a_k^2 < \infty$ .

Ainsi, d'après le premier point du théorème 2.2.4, la limite  $\sum_{k=1}^{\infty} d_k$  existe, on la notera  $W$ . On a donc  $a_n S_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \Gamma(2p)L$ , et  $a_n(S_n - \mathbb{E}(S_n)) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} W$ . Or on peut déduire de (4) que pour  $n \geq 1$ ,  $a_n \mathbb{E}(S_n) = 2q - 1$ . On déduit donc de tout cela la relation entre  $W$  et  $L$  :  $W = \Gamma(2p)L - (2q - 1)$ .

Vérifions maintenant les 2 conditions supplémentaires du théorème. Tout d'abord on remarque que pour  $k \geq 1$ , on a  $d_k = a_k \varepsilon_k$ , et donc d'après (6) et (1),

$$\mathbb{E}(d_k^2 | \mathcal{F}_{k-1}) = a_k^2 \left( 1 - \left( \frac{S_{k-1}}{k-1} \right)^2 (2p-1)^2 \right) \sim a_k^2 \sim \frac{\Gamma(2p)^2}{k^{4p-2}} \quad \text{p.s.}$$

Ainsi par comparaison série-intégrale, on obtient

$$V_n^2 := \sum_{k=n}^{\infty} \mathbb{E}(d_k^2 | \mathcal{F}_{k-1}) \sim \sum_{k=n}^{\infty} \frac{\Gamma(2p)^2}{k^{4p-2}} \sim \frac{\Gamma(2p)^2}{(4p-3)n^{4p-3}} \sim \frac{n a_n^2}{(4p-3)} \quad \text{p.s.}$$

Ainsi

$$\frac{V_n^2}{\frac{n a_n^2}{(4p-3)}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1,$$

cette quantité étant bornée, on a donc une convergence  $L^1$ , donc

$$\mathbb{E} \left( \frac{V_n^2}{\frac{na_n^2}{(4p-3)}} \right) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1, \quad \text{i.e.} \quad \frac{\mathbb{E}(V_n^2)}{\frac{na_n^2}{(4p-3)}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1.$$

Et on obtient alors

$$s_n^2 = \sum_{k=n}^{\infty} \mathbb{E}(d_k^2) \sim \frac{na_n^2}{(4p-3)},$$

ce qui implique

$$\frac{V_n^2}{s_n^2} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 1,$$

ce qui permet de vérifier la première condition. Vérifions maintenant la deuxième : soit  $\varepsilon > 0$ , alors

$$\frac{1}{s_n^2} \sum_{k=n}^{\infty} \mathbb{E}(d_k^2 \mathbf{1}_{\{|d_k| > \varepsilon s_n\}}) \leq \frac{1}{\varepsilon^2 s_n^4} \sum_{k=n}^{\infty} \mathbb{E}(d_k^4) \leq \frac{16}{\varepsilon^2 s_n^4} \sum_{k=n}^{\infty} a_k^4 \quad \text{p.s.}$$

Or

$$\frac{1}{s_n^4} \sum_{k=n}^{\infty} a_k^4 \sim \frac{(4p-3)^2}{n^2 a_n^4} \times \frac{\Gamma(2p)^4}{(8p-5)n^{8p-5}} \sim \frac{C}{n},$$

avec  $C \in \mathbb{R}$ , et donc

$$\frac{1}{s_n^2} \sum_{k=n}^{\infty} \mathbb{E}(d_k^2 \mathbf{1}_{\{|d_k| > \varepsilon s_n\}}) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

Ainsi, on déduit grâce au théorème :

$$\frac{W - \tilde{M}_n}{s_{n+1}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1),$$

et donc, comme  $W = \Gamma(2p)L - (2q-1)$ ,  $\tilde{M}_n = a_n(S_n - \mathbb{E}(S_n))$ , et  $a_n \mathbb{E}(S_n) = (2q-1)$ , on obtient

$$\frac{\Gamma(2p)L - a_n S_n}{a_n \sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4p-3}\right),$$

et ainsi, grâce à (1), on déduit

$$\frac{n^{2p-1}L - S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{4p-3}\right).$$

□

### 3 Utilisation des arbres aléatoires récursifs

Le rapprochement entre la marche aléatoire de l'éléphant et les arbres aléatoires récursifs a été fait pour la première fois dans [Kürsten, 2016] afin d'utiliser les résultats déjà obtenus sur la marche aléatoire pour obtenir des propriétés sur les arbres. Puis ce lien a été développé dans [Businger, 2018] en se servant des résultats obtenus sur les arbres dans [Baur and Bertoin, 2015] pour en déduire le comportement asymptotique d'une variante de la marche aléatoire de l'éléphant.

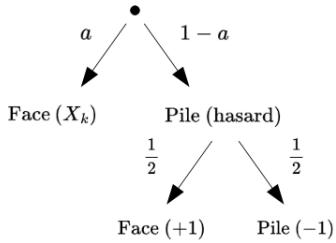
Nous allons ici utiliser les résultats de [Baur and Bertoin, 2015] afin d'obtenir plus d'informations sur le comportement asymptotique de notre marche dans le cas super-diffusif, et plus particulièrement, pour en savoir plus sur la variable limite  $L$ . Les résultats utilisés étaient exposés dans les articles de façon très théorique, et tout ce qui suit résulte d'un travail de reformulation et d'adaptation de ces modèles et résultats afin de les appliquer dans notre cadre.

Dans toute cette partie, nous allons supposer  $p > \frac{1}{2}$ .

### 3.1 Définition des arbres aléatoires récursifs et lien avec la marche de l'éléphant

Pour voir le lien entre notre marche aléatoire et les arbres récursifs, il est tout d'abord nécessaire d'introduire une formulation équivalente de la marche aléatoire de l'éléphant.

**Lemme 3.1.1.** *La formulation de la marche aléatoire de l'éléphant donnée en introduction est équivalente à la formulation suivante. L'éléphant part toujours de  $S_0 = 0$ , et son premier pas reste une loi de Rademacher de paramètre  $q$ . A l'étape  $n$ , l'éléphant se rappelle toujours d'une étape  $k \in \llbracket 1, n-1 \rrbracket$  parmi celles qu'il a déjà faites, mais cette fois ci, afin de décider le pas qu'il fera, il va lancer 2 pièces. La première pièce donne Face avec une probabilité  $a := 2p - 1$ , et Pile avec une probabilité  $1 - a$ . Si l'éléphant obtient Face, il refait exactement le pas de l'étape  $k$ . S'il obtient Pile, il lance la deuxième pièce, qui donnera Face avec probabilité  $\frac{1}{2}$ , et Pile avec probabilité  $\frac{1}{2}$ , Face correspondant à un pas à droite et Pile à un pas à Gauche. Ainsi, s'il fait Pile à la première pièce, il va alors uniformément au hasard à gauche ou à droite.*



*Démonstration.* Pour vérifier que cette formulation est effectivement équivalente à celle énoncée au début, nous allons calculer, pour cette formulation, la loi conditionnelle de  $X_{n+1}$  sachant  $X_1, \dots, X_n$  pour  $n \geq 1$  et vérifier qu'elle est bien égale à celle calculée précédemment pour la première formulation. Soit  $x \in \{-1, 1\}$ . On rappelle que  $\beta_{n+1}$  désigne la variable suivant une loi uniforme dans  $\llbracket 1, n \rrbracket$  représentant l'étape passée dont on se souvient à l'étape  $n+1$ . On a alors

$$\begin{aligned}
 \mathbb{P}(X_{n+1} = x | X_1, \dots, X_n) &= \sum_{k=1}^n \mathbb{P}_{X_1, \dots, X_n}(\beta_{n+1} = k) \times \mathbb{P}_{X_1, \dots, X_n}(X_{n+1} = x | \beta_{n+1} = k) \quad \text{p.s.} \\
 &= \sum_{k=1}^n \frac{1}{n} \left( \mathbf{1}_{\{X_k=x\}} \left( a + \frac{1}{2}(1-a) \right) + \mathbf{1}_{\{X_k=-x\}} \frac{1}{2}(1-a) \right) \quad \text{p.s.} \\
 &= \frac{1}{n} \sum_{k=1}^n (\mathbf{1}_{\{X_k=x\}} p + \mathbf{1}_{\{X_k=-x\}} (1-p)) \quad \text{p.s.}
 \end{aligned}$$

Et on vérifie que cela est bien égale à la loi obtenue pour la première formulation, cf équation (2).  $\square$

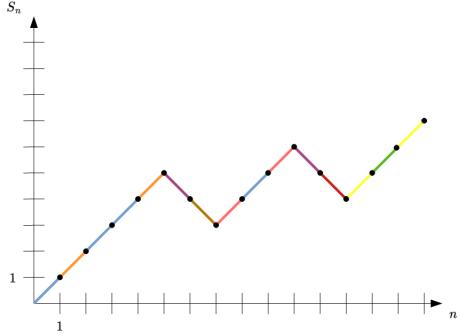
Cette formulation va nous permettre de faire facilement le lien avec les arbres aléatoires récursifs dont voici la définition.

**Définition 3.1.2.** *Un arbre aléatoire récursif est un arbre construit de la façon suivante : la racine est numérotée à 1, puis pour chaque nouveau noeud  $i \geq 2$ , on choisit uniformément au hasard le noeud auquel on le rattache parmi ceux déjà existants, donc uniformément au hasard parmi  $\llbracket 1, i-1 \rrbracket$ . Les noeuds d'un arbre aléatoire récursif de taille  $n$  sont donc numérotés de 1 à  $n$ , et pour chaque noeud, on appellera étiquette le numéro qui lui est attribué. On appelle percolation de Bernoulli sur les arêtes le processus suivant : chaque arête de l'arbre est conservée avec une certaine probabilité  $s$ , pour  $s \in [0, 1]$  et supprimée avec une probabilité  $1-s$ .*

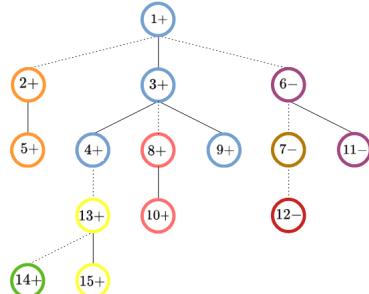
On voit donc apparaître le lien entre notre marche et ces arbres : nous allons construire une "forêt de mémoire" pour la marche, et cette forêt sera un arbre aléatoire récursif sur lequel on pratique la percolation de

Bernoulli. On crée un premier noeud pour le premier pas, qui sera donc la racine numérotée 1, à laquelle on associe un signe +1 ou -1 en fonction du pas effectué. Puis pour chaque nouveau pas  $n$ , on rattache le  $n$ -ième noeud au  $k$ -ième noeud, où  $k$  est l'étape dont on se souvient. Si la première pièce fait Face, on conserve l'arête (avec donc une probabilité  $a$ ), et on associe au  $n$ -ième noeud le même signe que celui du  $k$ -ième noeud. Si la première pièce fait Pile, on supprime l'arête que l'on vient de créer, et le signe du  $n$ -ième noeud dépend du pas obtenu à l'aide de la deuxième pièce.

**Exemple 3.1.1.** Illustrons cela avec une trajectoire de marche aléatoire de l'éléphant et la forêt de mémoire qui lui correspond.



(a) Trajectoire de la marche aléatoire



(b) Forêt de mémoire associée

**Remarque 3.1.3.** On remarque que dans chaque composante de l'arbre, tous les noeuds ont le même signe. Pour la composante contenant 1, ce signe est déterminé par la Rademacher de paramètre  $q$  qu'est le premier pas. Chacune des autres composantes se crée grâce à un lancer de la première pièce donnant pile, et pour chacune, le signe est alors déterminé par le lancer de la deuxième pièce. Ainsi le signe de la première composante est une Rademacher de paramètre  $q$ , et le signe de chacune des autres composantes est une Rademacher de paramètre  $\frac{1}{2}$ .

**Définition 3.1.4.** On appellera "cluster" chaque composante de l'arbre auquel on a appliqué la percolation de Bernoulli. Pour tout  $i \geq 1$ , on note  $c_{i,n}$  le cluster à l'étape  $n$  dont la racine a pour étiquette  $i$ , et  $|c_{i,n}|$  son cardinal. Si  $i$  n'est pas racine d'un cluster, alors  $c_{i,n} = \emptyset$  et  $|c_{i,n}| = 0$ . Pour chaque cluster  $c_{i,n}$ , on note  $m_i \in \{+1, -1\}$  le signe associé. D'après la remarque précédente,  $m_1$  suit une Rademacher de paramètre  $q$ , et pour tout  $i \geq 2$ ,  $m_i$  suit une Rademacher de paramètre  $\frac{1}{2}$ . On voit que pour tout  $n \geq 1$ , on peut alors écrire  $S_n$  de la façon suivante :

$$S_n = \sum_{i=1}^n |c_{i,n}| m_i.$$

De plus, pour tout  $i \geq 1$ , on appellera  $c'_{i,n}$  le sous cluster rattaché à  $i$  à l'étape  $n$ , c'est à dire tous les noeuds toujours attachés à  $i$  après percolation, sans que  $i$  ne soit nécessairement une racine. Ainsi pour tout  $i \geq 1$ ,  $|c'_{i,n}| \geq 1$ . Illustrons cela par un exemple.

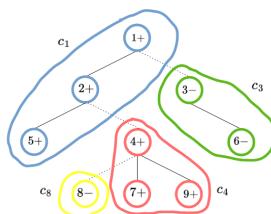


FIGURE 10 – Représentation des  $c_i$

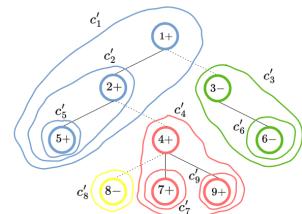


FIGURE 11 – Représentation des  $c'_i$

Ainsi, on voit donc comment les résultats existants sur les arbres aléatoires récursifs, et en particulier sur la taille des clusters, vont pouvoir nous aider à obtenir des résultats sur la marche aléatoire.

### 3.2 Utilisation des arbres pour approfondir le cas super-diffusif

Rappelons tout d'abord les résultats obtenus par [Baur and Bertoin, 2015] sur la taille des clusters. Les résultats suivants correspondent au Théorème 3.1 et au Lemme 3.3 de l'article.

**Théorème 3.2.1.** *[Admis] Pour tout  $0 \leq a \leq 1$ , on a*

$$\frac{|c_{1,n}|}{n^a} \xrightarrow[n \rightarrow \infty]{p.s.} X_1,$$

où la variable  $X_1$  a une distribution de Mittag-Leffler de paramètre  $a$  donnée par :

$$f(x) = \frac{1}{a} \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{k!} \Gamma(ka+1) x^{k-1} \sin(\pi ka).$$

Plus simplement, sa fonction génératrice des moments s'exprime de la façon suivante : pour  $t \in \mathbb{R}$ ,

$$M_{X_1}(t) = \sum_{k=0}^{+\infty} \frac{\Gamma(k+1)}{\Gamma(ak+1)} \frac{t^k}{k!}.$$

**Corollaire 3.2.2.** *[Admis] Pour tout  $0 \leq a \leq 1$ , et pour tout  $i \geq 2$ , on a*

$$\frac{|c'_{i,n}|}{n^a} \xrightarrow[n \rightarrow \infty]{p.s.} \rho_i,$$

où  $\rho_i$  a la même loi que  $\beta_i^a X_1$ , et  $\beta_i$  est une variable de loi Beta de paramètre  $(1, i-1)$  indépendante de  $X_1$ .

Mais comme on le voit sur la figure 11, les  $c'_{i,n}$  comptent avec beaucoup de répétitions les mêmes noeuds. Si l'on veut utiliser le corollaire précédent, il est donc nécessaire de ne compter que les  $c'_{i,n}$  des indices étant la racine de leur cluster.

**Définition 3.2.3.** *Pour  $k \geq 1$ , on définit  $i_k$  l'indice étant à la racine du  $k$ -ième cluster, en les numérotant par ordre d'apparition. Ainsi, pour tout  $n \geq 1$ , on peut écrire*

$$S_n = |c_{1,n}| m_1 + \sum_{k \geq 2} |c'_{i_k,n}| m_k, \quad (16)$$

où  $m_k$  désigne le signe du  $k$ -ième cluster. Il correspond à  $m_{i_k}$  dans les notations de la définition 3.1.4, mais cette notation est plus lourde et laisse sous entendre que ce signe dépend de  $i_k$  alors que ce n'est pas le cas.

Si l'on reprend la formulation donnée dans le lemme 3.1.1, on voit que chaque création d'un nouveau cluster correspond au fait que la première pièce fait Pile, et ce avec probabilité  $1-a$ . Ainsi, en se souvenant que la loi géométrique est la loi du premier succès dans un schéma de Bernoulli, on en déduit que  $i_1 = 1$  et que pour tout  $k \geq 1$ ,  $i_{k+1}$  a la même loi que  $i_k + G_k$  où  $G_k$  suit une loi géométrique de paramètre  $1-a$ , les  $(G_k)_{k \geq 1}$  étant tous indépendants. Par conséquent, pour  $k \geq 2$ ,  $i_k$  a la même loi que  $1 + B_k$  où  $B_k$  suit une loi binomiale négative de paramètres  $(1-a, k-1)$ , qui compte le nombre d'essais avant d'obtenir  $k-1$  succès dans un schéma de Bernoulli où le succès a une probabilité  $1-a$ .

**Exemple 3.2.1.** Dans l'exemple présenté sur la figure 11, on a donc  $i_1 = 1$ ,  $i_2 = 3$ ,  $i_3 = 4$  et  $i_4 = 8$ .

**Remarque 3.2.4.** *En fait, les indices  $i_k$  dépendent de  $n$  : pour  $k$  supérieur au nombre de cluster à l'étape  $n$ , on considère  $i_k = 0$ . Mais nous n'en tiendrons pas compte dans nos notations, tout d'abord pour ne pas alourdir les résultats à venir, et ensuite parce que lorsque  $n$  tend vers l'infini, le nombre de cluster à l'étape  $n$  tend vers l'infini presque sûrement. En effet le nombre de cluster à l'étape  $n$  a la même loi que  $1+b_n$  où  $b_n$  suit une variable binomiale de paramètres  $(n-1, 1-a)$  : il correspond au nombre de fois où la première pièce a fait Pile sur les  $n-1$  lancers effectués, +1 pour le cluster initial. Or  $b_n \xrightarrow[n \rightarrow \infty]{p.s.} \infty$ , et il en est donc de même pour le nombre de cluster, et donc lorsque  $n$  tend vers l'infini, les  $i_k$  sont presque sûrement tous non nuls.*

On a maintenant à disposition tous les ingrédients nécessaires afin d'obtenir l'expression de la variable limite  $L$  présente dans les théorèmes limites du cas super-diffusif.

**Théorème 3.2.5.** *Reprendons les notations introduites dans la définition 3.1.4, le théorème 3.2.1, le corollaire 3.2.2 et la définition 3.2.3. Pour  $a > \frac{1}{2}$ , c'est à dire pour  $p > \frac{3}{4}$ , la série  $\sum_{k \geq 2} \beta_{i_k}^a m_k$  converge, et alors on peut exprimer la variable limite  $L$  de la façon suivante :*

$$L = X_1 m_1 + \sum_{k \geq 2} \rho_{i_k} m_k.$$

**Remarque 3.2.6.** *Grâce au théorème 2.5.1, on sait en fait déjà que cette série converge, puisque l'on a obtenu  $\frac{S_n}{n^a} \xrightarrow[n \rightarrow \infty]{p.s.} L$ , en sachant déjà que  $L$  est une variable finie. Mais nous allons ici prouver la convergence de cette série d'une autre façon.*

Afin de prouver notre théorème, nous allons utiliser le théorème des trois séries de Kolmogorov dans le cas de variables non indépendantes, que l'on peut retrouver dans [Brown, 1971], voir théorème 1.

**Théorème 3.2.7. [Admis]** *Soit  $(Y_n)_{n \geq 1}$  une suite de variables aléatoires, et soit  $(\mathcal{G}_n)_{n \geq 1}$  la filtration naturelle définie par  $\mathcal{G}_n = \sigma(Y_1, \dots, Y_n)$ . Soit  $K < \infty$  une constante arbitraire positive. Pour  $n \geq 1$ , on note  $Z_n = Y_n \mathbf{1}_{\{|Y_n| \leq K\}}$ . Alors  $\sum_n Z_n$  converge si et seulement si*

1.  $\sum_j \mathbb{P}(|Y_j| > K | \mathcal{G}_{j-1}) < \infty$ ,
2.  $\sum_j \mathbb{E}(Z_j | \mathcal{G}_{j-1})$  converge,
3.  $\sum_j \mathbb{E}\left((Z_j - \mathbb{E}(Z_j | \mathcal{G}_{j-1}))^2 | \mathcal{G}_{j-1}\right) < \infty$ .

Utilisons ceci afin de prouver le théorème 3.2.5.

*Démonstration.* Tout d'abord, nous allons utiliser le théorème 3.2.7 afin de montrer la convergence de la série  $\sum_{k \geq 2} \beta_{i_k}^a m_k$ . On pose pour tout  $k \geq 2$ ,  $Y_k = \beta_{i_k}^a m_k$ , et on note  $\mathcal{G}_k = \sigma(Y_2, \dots, Y_k)$ . Alors comme  $|m_k| = 1$ , et comme la loi Beta est à support dans  $[0, 1]$ , on sait que  $|Y_k| \leq 1$  presque sûrement. Donc

$$\forall k \geq 2, \quad \mathbb{P}(|Y_k| > 1 | \mathcal{G}_{k-1}) = 0,$$

et donc la première condition du théorème est vérifiée, pour  $K = 1$ . On déduit également de cela que pour tout  $k \geq 2$ ,  $Y_k = Z_k$  presque sûrement et donc,  $\mathbb{E}(Z_k | \mathcal{G}_{k-1}) = \mathbb{E}(Y_k | \mathcal{G}_{k-1})$  p.s. Or  $Y_k = \beta_{i_k}^a m_k$ , et  $m_k$  est indépendante de  $\beta_{i_k}^a$  et de  $Y_2, \dots, Y_{k-1}$ , donc

$$\mathbb{E}(Y_k | \mathcal{G}_{k-1}) = \mathbb{E}(\beta_{i_k}^a | \mathcal{G}_{k-1}) \times \mathbb{E}(m_k) = 0 \quad \text{p.s.}$$

Et donc la seconde condition est également vérifiée. Passons maintenant à la dernière condition. On a

$$\mathbb{E}\left((Z_k - \mathbb{E}(Z_k | \mathcal{G}_{k-1}))^2 | \mathcal{G}_{k-1}\right) = \mathbb{E}\left((Y_k - \mathbb{E}(Y_k | \mathcal{G}_{k-1}))^2 | \mathcal{G}_{k-1}\right) = \mathbb{E}(Y_k^2 | \mathcal{G}_{k-1}) = \mathbb{E}(\beta_{i_k}^{2a} | \mathcal{G}_{k-1}) \quad \text{p.s.}$$

On pose pour  $k \geq 2$ ,  $\mathcal{H}_k = \sigma(i_2, \dots, i_k, Y_2, \dots, Y_k)$ , alors  $\mathcal{G}_{k-1} \subset \mathcal{H}_k$ , et donc

$$\mathbb{E}(\beta_{i_k}^{2a} | \mathcal{G}_{k-1}) = \mathbb{E}(\mathbb{E}(\beta_{i_k}^{2a} | \mathcal{H}_k) | \mathcal{G}_{k-1}) \quad \text{p.s.}$$

Or le moment d'ordre  $j$  d'une loi Beta de paramètre  $(\alpha, \beta)$  est  $\frac{B(\alpha+j, \beta)}{B(\alpha, \beta)}$ , où  $B$  désigne la fonction Beta. Cela implique donc

$$\mathbb{E}(\beta_{i_k}^{2a} | \mathcal{H}_k) = \frac{B(1+2a, i_k-1)}{B(1, i_k-1)} = \frac{\Gamma(1+2a)\Gamma(i_k)}{\Gamma(2a+i_k)} \quad \text{p.s.}$$

Or pour tout  $k \geq 1$ ,  $i_k \geq k$  presque sûrement, donc  $i_k$  tend presque sûrement vers l'infini. Ainsi en utilisant l'équivalent de la fonction Gamma suivant :  $\Gamma(x+1) \sim \left(\frac{x}{e}\right)^x \sqrt{2\pi x}$ , on obtient :

$$\frac{\Gamma(1+2a)\Gamma(i_k)}{\Gamma(2a+i_k)} \sim \frac{\Gamma(1+2a)}{i_k^{2a}} \quad \text{p.s.}$$

Et comme pour tout  $k \geq 1$ ,  $i_k > k+1$  p.s.,

$$\frac{\Gamma(1+2a)}{i_k^{2a}} < \frac{\Gamma(1+2a)}{(k+1)^{2a}} \quad \text{p.s.}$$

On note  $a_k = \mathbb{E}(\beta_{i_k}^{2a} | \mathcal{H}_k)$ ,  $b_k = \frac{\Gamma(1+2a)}{i_k^{2a}}$  et  $c_k = \frac{\Gamma(1+2a)}{(k+1)^{2a}}$ . Comme  $a_k \sim b_k$ , on sait que  $\frac{a_k}{b_k} \xrightarrow[k \rightarrow \infty]{\text{p.s.}} 1$ . Ainsi,

$$\forall \ell \geq 1, \quad \exists N_\ell \geq 1, \quad \forall k \geq N_\ell, \quad b_k \left(1 - \frac{1}{\ell}\right) \leq a_k \leq b_k \left(1 + \frac{1}{\ell}\right) \quad \text{p.s.}$$

Donc à partir d'un certain rang,

$$a_k \leq c_k \quad \text{p.s.}, \quad \text{c'est à dire dans notre contexte} \quad \mathbb{E}(\beta_{i_k}^{2a} | \mathcal{H}_k) \leq \frac{\Gamma(1+2a)}{(k+1)^{2a}} \quad \text{p.s.}$$

On peut passer à l'espérance conditionnelle selon  $\mathcal{G}_{k-1}$  dans cette inégalité et on obtient :

$$\mathbb{E}(\beta_{i_k}^{2a} | \mathcal{G}_{k-1}) \leq \frac{\Gamma(1+2a)}{(k+1)^{2a}} \quad \text{p.s.}$$

Ainsi, si  $a > \frac{1}{2}$ , on vérifie la troisième et dernière condition du théorème 3.2.7, qui nous donne la convergence de la série  $\sum_{k \geq 2} \beta_{i_k}^a m_k$ .

Passons à la suite du théorème. On sait d'après le théorème 2.5.1 que d'une part  $\frac{S_n}{n^a} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} L$ . D'autre part, on voit grâce à l'équation (16) :

$$\frac{S_n}{n^a} = \frac{|c_{1,n}|}{n^a} m_1 + \sum_{k \geq 2} \frac{c'_{i_k,n}}{n^a} m_k.$$

Et donc on déduit grâce au théorème 3.2.1 et au corollaire 3.2.2 :

$$\frac{S_n}{n^a} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X_1 m_1 + \sum_{k \geq 2} \rho_{i_k} m_k.$$

On vient de vérifier que le terme limite est bien défini et fini, et donc on obtient le résultat souhaité par unicité de la limite.

**Remarque 3.2.8.** Si  $a < \frac{1}{2}$ , ce qui correspond au cas diffusif, on a vu  $\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{3-4p}\right)$ , et donc  $\frac{S_n}{n^a}$  ne peut pas converger presque sûrement vers une variable finie. On en déduit que la série étudiée dans la preuve précédente diverge dans ce cas. De même pour le cas critique correspondant à  $a = \frac{1}{2}$ , on a vu  $\frac{S_n}{\sqrt{n \log(n)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$ , donc  $\frac{S_n}{n^a}$  ne peut pas converger presque sûrement vers une variable finie.

□

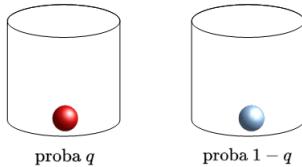
## 4 Les urnes de Polya

Le rapprochement entre la marche aléatoire de l'éléphant et les urnes de Polya a été fait pour la première fois par [Baur and Bertoin, 2016]. Le lien entre les deux se fait très facilement, et cette modélisation peut permettre d'étudier d'autres variantes de la marche de l'éléphant en faisant seulement varier les conditions de remplacement dans l'urne.

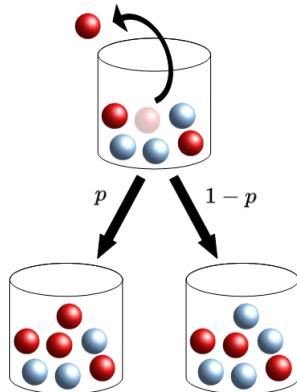
## 4.1 Lien avec les urnes

Tout d'abord, commençons par définir l'urne de Polya que nous allons utiliser par la suite.

**Définition 4.1.1.** On définit  $(U_n)_{n \geq 0}$  une urne avec des boules de couleurs rouge et bleue. La composition de l'urne à l'étape  $n$  est donnée par  $U_n = (R_n, B_n)$  pour  $n \geq 0$ , où  $R_n$  désigne le nombre de boules rouges et  $B_n$  désigne le nombre de boules bleues. A l'étape 1, la composition est donnée par  $(1, 0)$  avec probabilité  $q$ , et  $(0, 1)$  avec probabilité  $1 - q$ .



Puis à l'étape  $n$ , pour  $n \geq 2$ , on tire au hasard une boule, on regarde sa couleur et on la remet. Ensuite, avec probabilité  $p$ , on rajoute une boule de la même couleur, et avec probabilité  $1 - p$ , on rajoute une boule de la couleur opposée.



Au vu de cette définition, il semble que les boules rouges vont désigner les pas à droite, et les boules bleues les pas à gauche, et c'est effectivement ce que l'on va vérifier ici.

**Proposition 4.1.2.** Pour une marche aléatoire de l'éléphant partant de  $S_0 = 0$  et telle que  $S_1 = R_1 - B_1$ , alors

$$\forall n \in \mathbb{N}, \quad S_n \stackrel{\mathcal{L}}{=} R_n - B_n.$$

*Démonstration.* Tout d'abord, notons pour tout  $k \geq 1$ ,  $Y_k$  qui vaut  $+1$  si la boule rajoutée dans l'urne à l'étape  $k$  est rouge, et  $-1$  si la boule ajoutée est bleue. Alors pour tout  $n \geq 1$ ,  $R_n - B_n = \sum_{k=1}^n Y_k$ . Pour vérifier l'égalité en loi de notre proposition, nous allons calculer la loi conditionnelle de  $Y_{n+1}$  sachant  $Y_1, \dots, Y_n$  pour  $n \geq 1$  et vérifier qu'elle est bien égale à celle de  $X_{n+1}$  sachant  $X_1, \dots, X_n$  calculée pour la marche de l'éléphant. Soit  $y \in \{-1, 1\}$ . On note  $\tau_{n+1}$  le numéro de la boule piochée à l'étape  $n + 1$ , en considérant qu'on numérote chaque boule en fonction de l'étape à laquelle elle a été ajoutée. Alors on a

$$\begin{aligned} \mathbb{P}(Y_{n+1} = y | Y_1, \dots, Y_n) &= \sum_{k=1}^n \mathbb{P}_{Y_1, \dots, Y_n}(\tau_{n+1} = k) \times \mathbb{P}_{Y_1, \dots, Y_n}(Y_{n+1} = y | \tau_{n+1} = k) \quad \text{p.s.} \\ &= \frac{1}{n} \sum_{k=1}^n (\mathbb{1}_{\{Y_k=y\}} p + \mathbb{1}_{\{Y_k=-y\}} (1-p)) \quad \text{p.s.} \end{aligned}$$

Et on retrouve donc la loi calculée pour la marche de l'éléphant obtenue à l'équation (2), ce qui prouve l'égalité en loi souhaitée.  $\square$

On voit donc que le lien se fait assez directement et cette modélisation peut s'adapter facilement à d'autres alternatives de la marche.

## 4.2 Utilisation de cette approche

Nous n'allons en fait pas détailler les résultats que l'on peut obtenir grâce à cette méthode. En effet les théorèmes de convergence liés aux urnes de Polya traitent de convergences de processus qui dépendent du temps, et ces notions sortent du cadre fixé pour le stage. Cependant, sans rentrer dans les détails des convergences et de leur preuve, voyons quelques points importants liés à cette méthode.

Tout d'abord les théorèmes obtenus grâce à l'étude des urnes de Polya impliquent les théorèmes que nous avons obtenus à l'aide des martingales. Cependant, si les urnes de Polya permettent d'obtenir une version plus poussée des convergences étudiées, les martingales permettent d'obtenir d'autres types de théorèmes que nous n'étudierons pas.

Ensuite, nous pouvons facilement voir grâce aux urnes de Polya la démarcation entre les trois régimes. En effet, le critère déterminant afin d'étudier une urne de Polya est l'espérance de sa matrice de remplacement, qui détermine de quelle façon on ajoute des boules à chaque étape. Le coefficient à la ligne  $i$  et à la colonne  $j$  de cette matrice détermine le nombres de boules de couleur  $j$  que l'on ajoute si l'on a tiré une boule de couleurs  $i$ . Ici, pour l'urne étudiée, la matrice de remplacement est l'identité avec probabilité  $p$ , et la matrice  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  avec probabilité  $1 - p$ , et son espérance est donc donnée par

$$\begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

qui a comme valeurs propres  $\lambda_1 = 1$  et  $\lambda_2 = 2p - 1$ , avec donc  $\lambda_1 \geq \lambda_2$ . Et les théorèmes de la référence [Janson, 2004] utilisés par [Baur and Bertoin, 2016] nous indiquent que les théorèmes limites dépendent de la position de  $\frac{\lambda_2}{\lambda_1}$  par rapport à  $\frac{1}{2}$ . Ici, cela dépend donc de la position de  $2p - 1$  par rapport à  $\frac{1}{2}$ , et donc de la position de  $p$  par rapport à  $\frac{3}{4}$ , ce qui nous ramène à la disjonction habituelle des trois régimes.

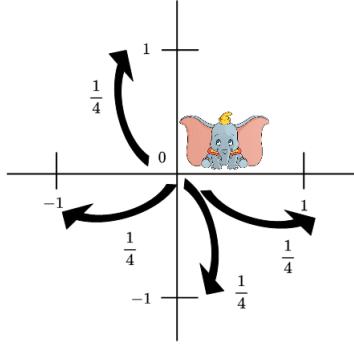
Enfin, cette méthode peut être adaptée à d'autres marches aléatoires en modifiant la matrice de remplacement, mais nous verrons cela par la suite.

## 5 La marche de l'éléphant en dimension supérieure

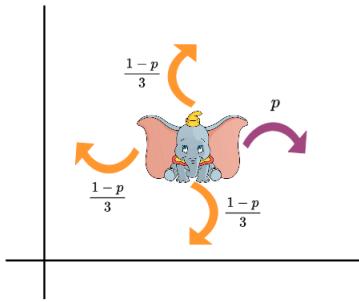
Pour complexifier la marche de l'éléphant, on peut tout d'abord essayer de voir comment l'étendre à la dimension supérieure, et si les différents outils développés précédemment sont toujours utilisables.

### 5.1 Définition de la marche en dimension supérieure

Soit  $d \geq 2$  la dimension de l'espace. A l'instant initial, l'éléphant se place toujours à l'origine. Puis à l'étape 1, nous allons pour simplifier considérer le cas où l'éléphant choisit une des  $2d$  directions qui s'offrent à lui uniformément au hasard, avec donc une probabilité  $\frac{1}{2d}$  pour chacune. Illustrons cela pour la dimension 2.



Ensuite pour tout étape  $n \geq 2$ , l'éléphant va toujours se rappeler uniformément au hasard d'une étape  $k \in \llbracket 1, n-1 \rrbracket$  parmi toutes celles qu'il a déjà effectuées. Puis il va faire le même pas avec probabilité  $p$ , et sinon il va aller uniformément au hasard dans l'une des directions restantes, avec donc probabilité  $\frac{1-p}{2d-1}$  pour chacune. Illustrons cela de nouveau en dimension 2 : sur l'exemple ci dessous, l'éléphant se rappelle d'un pas vers la droite.



Pour tout  $k \geq 1$ , on note de nouveau  $X_k$  le pas effectué à l'étape  $k$ , qui est donc un vecteur de taille  $d$  contenant un unique  $+1$  ou  $-1$ , et on garde alors l'expression, pour  $n \geq 1$ ,

$$S_n = \sum_{k=1}^n X_k.$$

Le cas  $p = \frac{1}{2d}$  correspond à la marche aléatoire symétrique en dimension  $d$ .

Nous allons voir dans la suite comment adapter les méthodes utilisées précédemment à l'étude de notre marche en dimension supérieure. Les résultats précis concernant la dimension  $d$  sont à retrouver dans [Bercu and Laulin, 2019].

## 5.2 Les martingales

L'approche martingale s'utilise de la même façon que pour la dimension 1, et va permettre d'obtenir les mêmes résultats.

**Proposition 5.2.1.** *On pose, pour  $n \geq 1$ ,  $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ . On définit  $(M_n)_{n \geq 0}$  par  $M_n = a_n S_n$  pour  $n \geq 0$ , où  $a_0 = a_1 = 1$ , et pour  $n \geq 2$  :*

$$a_n := \prod_{k=1}^{n-1} \gamma_k^{-1} = \frac{\Gamma(a+1)\Gamma(n)}{\Gamma(a+n)} \sim \frac{\Gamma(a+1)}{n^a}, \quad \text{où } \gamma_k := 1 + \frac{a}{k},$$

où  $a := \frac{2dp-1}{2d-1}$ . Alors  $(M_n)_{n \geq 0}$  est une  $(\mathcal{F}_n)_{n \geq 0}$ -martingale.

*Démonstration.* On note  $(e_1, \dots, e_d)$  la base canonique de  $\mathbb{R}^d$ . Soit  $n \geq 1$ . On sait que  $X_{n+1}$  est à valeurs dans  $\{e_1, \dots, e_d, -e_1, \dots, -e_d\}$ . Calculons donc  $\mathbb{E}(X_{n+1} | \mathcal{F}_n)$ .

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = \sum_{i=1}^d e_i \mathbb{P}(X_{n+1} = e_i | X_1, \dots, X_n) - \sum_{i=1}^d e_i \mathbb{P}(X_{n+1} = -e_i | X_1, \dots, X_n) \quad \text{p.s.}$$

Pour  $i \in \llbracket 1, d \rrbracket$ , notons  $N_n^+(i)$  le nombre de pas dans la direction  $+e_i$  faits jusqu'à l'étape  $n$ , et  $N_n^-(i)$  le nombre de pas dans la direction  $-e_i$  jusqu'à cette étape. Ainsi,

$$\begin{aligned} \mathbb{E}(X_{n+1} | \mathcal{F}_n) &= \sum_{i=1}^d e_i \left( \frac{N_n^+(i)}{n} \times p + \frac{n - N_n^+(i)}{n} \times \frac{1-p}{2d-1} \right) - \sum_{i=1}^d e_i \left( \frac{N_n^-(i)}{n} \times p + \frac{n - N_n^-(i)}{n} \times \frac{1-p}{2d-1} \right) \\ &= \frac{1}{n} \left( p - \frac{1-p}{2d-1} \right) \sum_{i=1}^d e_i (N_n^+(i) - N_n^-(i)). \end{aligned}$$

On remarque que  $S_n = \sum_{i=1}^d (N_n^+(i) - N_n^-(i))$ , et donc on en déduit

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = \left( \frac{2dp-1}{2d-1} \right) \frac{S_n}{n} = \frac{aS_n}{n}.$$

Cela implique donc

$$\mathbb{E}(S_{n+1} | \mathcal{F}_n) = S_n + \mathbb{E}(X_{n+1} | \mathcal{F}_n) = \left( 1 + \frac{a}{n} \right) S_n = \gamma_n S_n.$$

Et donc en posant  $M_n = a_n S_n$ , avec  $a_n = \prod_{k=1}^{n-1} \gamma_k^{-1}$ , on obtient bien

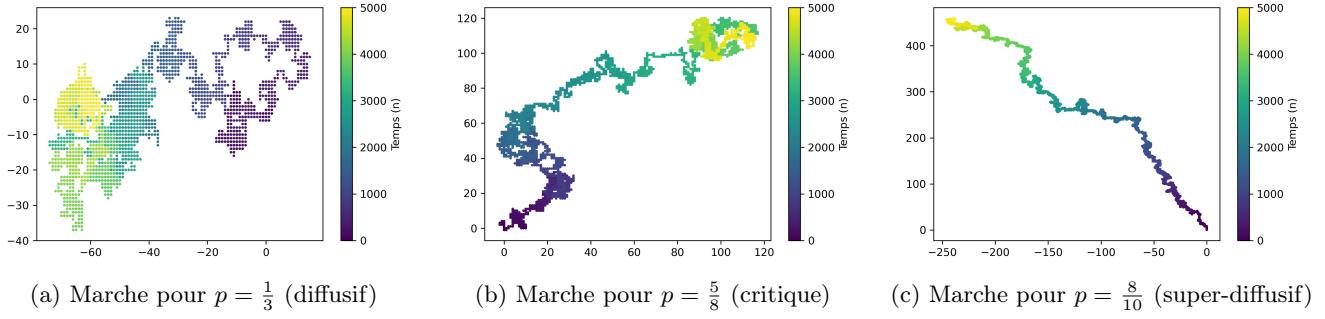
$$\mathbb{E}(M_{n+1} | \mathcal{F}_n) = M_n,$$

ce qui nous indique bien que  $(M_n)_{n \geq 0}$  est une martingale.  $\square$

A partir de là, on peut reprendre les mêmes idées que pour la dimension 1, même si comme on l'aperçoit dans la preuve précédente les calculs deviennent compliqués car contrairement à la dimension 1, savoir que l'on ne va pas dans une direction ne permet pas de déduire la direction que l'on a finalement choisie.

Mais les résultats restent les mêmes. En calculant la trace du compensateur, qui guide le comportement asymptotique, on peut voir qu'apparaît de nouveau la série  $\sum_k a_k^2$ , dont la convergence varie en fonction de la position de  $a$  par rapport à  $\frac{1}{2}$ , et donc de la position de  $p$  par rapport à  $p_d = \frac{2d+1}{4d}$ . A partir de cela, et de théorèmes sur les martingales similaires à ceux de la dimension 1, on déduit que le comportement asymptotique de la marche se divise toujours en trois régimes en fonction de  $p$  : le régime diffusif pour  $0 \leq p < p_d$ , le régime critique pour  $p = p_d$ , et le régime super-diffusif pour  $p_d < p \leq 1$ . Les théorèmes limites restent les mêmes pour chaque régime.

Ces trois régimes sont illustrés ci dessous pour la dimension 2 par des trajectoires de marche de l'éléphant après 5000 pas pour des valeurs de  $p$  dans chaque régime. On observe dans chaque cas la trajectoire sur le plan, et la couleur nous permet voir l'évolution dans le temps, le violet désignant le départ de la marche.



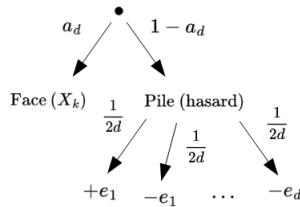
Ainsi, même si les calculs deviennent plus lourds, cette méthode s'adapte facilement à la dimension supérieure et permet d'obtenir exactement les mêmes résultats.

### 5.3 Les arbres aléatoires récursifs

Comme l'explique la partie précédente, les théorèmes limites en dimension  $d$  sont les mêmes que pour la dimension 1, cela implique donc la présence d'une variable limite  $L_d$  inconnue a priori dans le cas du régime super-diffusif. Mais, de même que pour la dimension 1, l'utilisation de la théorie sur les arbres aléatoires récursifs peut permettre de mieux connaître cette variable.

Afin d'utiliser les arbres aléatoires récursifs, on introduit une autre formulation de la marche similaire à celle de la dimension 1.

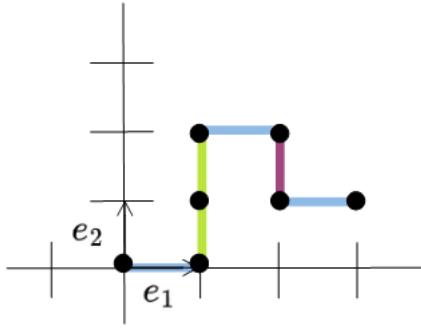
**Lemme 5.3.1.** *La formulation de la marche aléatoire de l'éléphant en dimension  $d$  énoncée précédemment est équivalente à la formulation suivante. L'éléphant part toujours de l'origine, et son premier pas est toujours choisi uniformément au hasard parmi les  $2d$  directions à disposition. A l'étape  $n$ , l'éléphant se rappelle toujours d'une étape  $k \in \llbracket 1, n-1 \rrbracket$  parmi celles qu'il a déjà faites, mais cette fois ci, afin de décider le pas qu'il fera, il va lancer une pièce et un dé. La pièce donne Face avec une probabilité  $a := \frac{2dp-1}{2d-1}$ , et Pile avec une probabilité  $1-a$ . Si l'éléphant obtient Face, il refait exactement le pas de l'étape  $k$ . S'il obtient Pile, il lance le dé, qui est équilibré et possède  $2d$  faces. Chaque face indique une direction, a donc une probabilité  $\frac{1}{2d}$  d'être obtenue, et la face effectivement obtenue déterminera la direction prise par l'éléphant.*



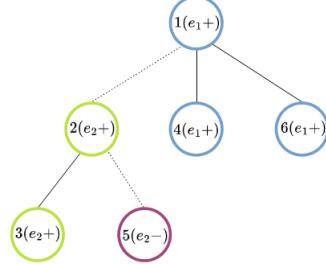
La preuve est la même qu'en dimension 1, et on vérifie bien que la loi est identique.

On peut alors réutiliser la même méthode de construction d'une forêt de mémoire, en adaptant seulement les étiquettes pour qu'elles contiennent la direction prise par l'éléphant à chaque étape. On procède donc de la façon suivante : à l'étape 1, on crée une racine numérotée 1, à laquelle on associe une direction, et donc un vecteur choisi uniformément au hasard parmi  $\{e_1, \dots, e_d, -e_1, \dots, -e_d\}$  selon le pas effectué. Puis pour chaque nouveau pas  $n$ , on rattache toujours le  $n$ -ième noeud au  $k$ -ième noeud, où  $k$  est l'étape dont on se souvient. Si la pièce fait Face, on conserve l'arête (avec probabilité  $a$ ), et on associe au  $n$ -ième noeud le même vecteur que celui du  $k$ -ième noeud. Si la pièce fait Pile, on supprime l'arête et on associe au noeud le vecteur déterminé par le dé.

**Exemple 5.3.1.** Illustrons cela avec une trajectoire de marche aléatoire de l'éléphant en dimension 2.



(a) Trajectoire de la marche aléatoire



(b) Forêt de mémoire associée

On peut alors reprendre les clusters définis précédemment, et faire le lien entre les arbres et la marche aléatoire.

**Définition 5.3.2.** *On reprend les notations introduites dans les définitions 3.1.4 et 3.2.3. Pour  $k \geq 1$ , on note donc toujours  $i_k$  l'indice à la racine du  $k$ -ième cluster, et  $|c'_{i_k,n}|$  sa taille. L'étiquette de ce cluster est composée d'un vecteur dans  $\{e_1, \dots, e_d\}$  et d'un signe dans  $\{+1, -1\}$ . Ainsi  $m_k$  désigne le signe du  $k$ -ième cluster, et donc  $m_k$  suit une loi de Rademacher de paramètre  $\frac{1}{2}$ . On notera  $V_k$  le vecteur du  $k$ -ième cluster, il suit donc une loi multinomiale à une seule épreuve où la probabilité de chaque résultat est  $\frac{1}{2d}$ , ce qui signifie que  $V_k$  prend une valeur uniformément au hasard parmi  $\{e_1, \dots, e_d\}$ , ce qui correspond bien à notre modèle. Et ainsi, pour  $n \geq 1$ , on peut exprimer  $S_n$  en fonction de toutes ces variables :*

$$S_n = \sum_{k \geq 1} |c'_{i_k,n}| m_k V_k.$$

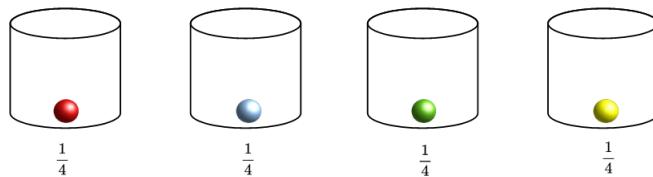
Ainsi, grâce au résultat des théorèmes 3.2.1 et 3.2.2, on obtient cette fois  $\frac{S_n}{n^{\alpha}} \xrightarrow[n \rightarrow \infty]{\text{P.s.}} \sum_{k \geq 1} \rho_{i_k} m_k V_k$ , ce qui nous

donne des informations sur la variable limite du cas super-diffusif, et on pourrait de nouveau vérifier que la série converge bel et bien pour  $p > p_d$ .

## 5.4 Les urnes de Polya

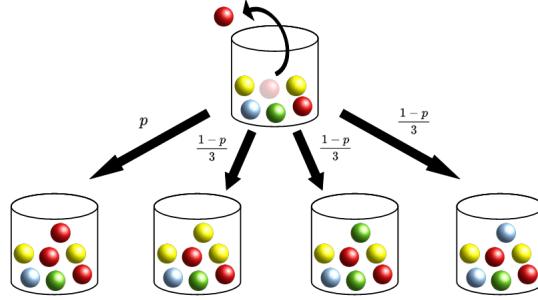
Tout comme les martingales et les arbres, les urnes de Polya s'adaptent très facilement en dimension  $d$ , et permettent d'obtenir le même type de résultats que pour la dimension 1. L'urne est définie de la même façon qu'en dimension 1 : en dimension 1, on a deux choix de direction, on met en conséquence 2 couleurs de boules, en dimension  $d$ , on a  $2d$  choix de direction, on met donc  $2d$  couleurs de boules.

**Définition 5.4.1.** *On définit  $(U_n)_{n \geq 0}$  une urne avec des boules de  $2d$  couleurs différentes. La composition de l'urne à l'étape  $n$  est donnée par  $U_n = (U_n^1, \dots, U_n^{2d})$  pour  $n \geq 0$ , où  $U_n^k$  désigne le nombre de boules de la couleur  $k$  à l'étape  $n$ . A l'étape 1, on met une boule d'une couleur que l'on choisit uniformément au hasard parmi les  $2d$  couleurs, avec probabilité  $\frac{1}{2d}$  pour chacune. Voici l'illustration de ceci en dimension 2.*



Puis à l'étape  $n$ , pour  $n \geq 2$ , on tire au hasard une boule, on regarde sa couleur et on la remet. Ensuite, avec probabilité  $p$ , on rajoute une boule de la même couleur, et sinon on rajoute une boule en choisissant sa couleur

uniformément parmi les couleurs restantes , avec donc probabilité  $\frac{1-p}{2d-1}$  pour chacune. De nouveau illustrons cela en dimension 2.



Chaque couleur est associée à une direction, et on associe donc les couleurs par pairs, chaque paire désignant une dimension de l'espace, et on retrouve alors notre marche.

**Proposition 5.4.2.** Pour une marche aléatoire de l'éléphant en dimension  $d$  partant de  $S_0 = 0_{\mathbb{R}^d}$ , et telle que  $S_1 = (U_1^1 - U_1^{d+1})e_1 + \dots + (U_1^d - U_1^{2d})e_d$ , alors

$$\forall n \in \mathbb{N}, \quad S_n \stackrel{\mathcal{L}}{=} (U_n^1 - U_n^{d+1})e_1 + \dots + (U_n^d - U_n^{2d})e_d.$$

La preuve se fait exactement de la même façon que pour la dimension 1. Ainsi, tous comme les deux autres étudiées précédemment, cette méthode s'adapte très facilement pour la dimension  $d$ .

De nouveau, nous n'allons pas rentrer dans les détails des résultats obtenus, qui peuvent être retrouvés dans [Bertenghi, 2022], mais nous pouvons tout de même observer comment cette méthode permet de voir apparaître les trois régimes. En effet, [Janson, 2004] indique qu'en dimension  $d$ , le critère déterminant est la position de  $\frac{\lambda_2}{\lambda_1}$  par rapport à  $\frac{1}{2}$ , où  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$  sont les valeurs propres de l'espérance de la matrice de remplacement, classées par ordre décroissant.

Ici la matrice de remplacement est donnée par l'identité avec probabilité  $p$ , et par  $J^k$  avec probabilité  $\frac{1-p}{2d-1}$  pour tout  $k$  dans  $\llbracket 1, 2d-1 \rrbracket$ , où

$$J = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

Son espérance est donc donnée par

$$\begin{bmatrix} p & \frac{1-p}{2d-1} & \cdots & \frac{1-p}{2d-1} \\ \frac{1-p}{2d-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{1-p}{2d-1} \\ \frac{1-p}{2d-1} & \cdots & \frac{1-p}{2d-1} & p \end{bmatrix}$$

et ses valeurs propres sont  $\lambda_1 = 1$  et  $\lambda_2 = \dots = \lambda_d = \frac{2dp-1}{2d-1}$ . On voit donc que les théorèmes limites dépendent de la position de  $\frac{2dp-1}{2d-1}$  par rapport à  $\frac{1}{2}$ , ce qui correspond à la position de  $p$  par rapport à  $p_d$ , et on retrouve bien les trois régimes.

Ainsi, les trois méthodes que nous avons développées pour étudier la dimension 1 sont faciles à étendre pour la dimension  $d$ , et permettent d'obtenir les mêmes résultats. Nous allons voir dans la suite si elles peuvent s'adapter à d'autres extensions de la marche de l'éléphant.

## 6 Ouvertures : extensions de la marche aléatoire de l'éléphant

Comme nous venons de le voir, les trois approches développées précédemment s'étendent très facilement à la dimension supérieure, mais sont-elles également applicables à d'autres extensions de la marche aléatoire de l'éléphant ? C'est ce que nous allons voir.

### 6.1 Quelques extensions déjà étudiées

Intéressons nous tout d'abord à quelques généralisations qui ont déjà été traitées.

Tout d'abord pour modifier la marche, on peut modifier la façon qu'a l'éléphant de se rappeler des événements passés : dans ce que nous avons vu précédemment, l'éléphant se rappelait uniformément au hasard d'un évènement dans le passé, mais il est possible de modifier cette distribution.

En effet, on peut supposer premièrement que l'éléphant possède une mémoire renforcée : plus on se souvient d'une étape passée, plus la probabilité qu'on se souvienne d'elle à l'avenir augmente. Cette version de la marche de l'éléphant a été traitée par [Laulin, 2022b], qui utilise pour cela l'approche martingale, même si cela se fait de façon beaucoup plus complexe que ce qui a été fait précédemment. La référence [Baur and Bertoin, 2016] donne également des pistes afin d'étudier cette extension à l'aide des urnes de Polya.

Toujours dans l'idée de modifier les probabilités de souvenir de chaque étape, on peut imaginer que l'on se souvient plus des événements récents que des plus anciens. C'est ce que l'on va appeler l'amnésie. Cette extension a été traitée par [Laulin, 2022a], qui utilise de nouveau l'approche martingale. On pourrait également imaginer l'inverse : les instants très proches du présent ont plus de poids que les instants très anciens.

Ensuite, au lieu de jouer sur les probabilités de mémoire, il est possible pour généraliser notre marche de modifier les types et le nombre de pas effectués à chaque étape. L'article [Businger, 2018] traite une extension de la sorte en utilisant les arbres aléatoires : si l'on reprend la formulation avec le lancer de pièces, on suppose que si l'éléphant fait Pile et donc fait un pas au hasard, il ne fait plus forcément un pas suivant une loi de Rademacher de paramètre  $\frac{1}{2}$ , mais suivant une autre loi qui pourrait être quelconque (même si dans l'article, des hypothèses sont faites sur cette loi).

En conclusion, ces trois exemples d'extension nous montrent que les approches développées précédemment peuvent également s'adapter pour des extensions de la marche de l'éléphant, même si elles deviennent parfois beaucoup plus complexes que ce qui a été fait pour la dimension 1 de la marche classique.

### 6.2 De nouvelles pistes...

Maintenant que nous avons vu ces différentes généralisations sur lesquels les approches peuvent fonctionner, nous pouvons nous interroger sur quelques autres extensions envisageables, et sur les méthodes pouvant être applicables pour les traiter.

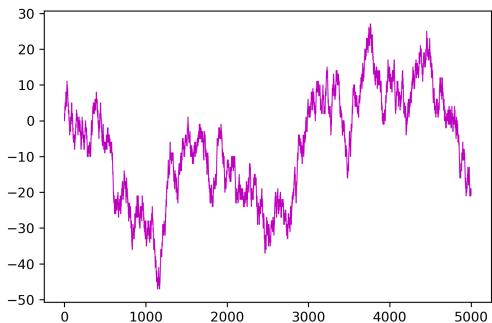
Premièrement, on pourrait imaginer que le paramètre de mémoire de la marche dépend du temps : plus le temps passe moins l'éléphant se rappelle de ce qu'il a fait précédemment. Cela se traduirait mathématiquement par un paramètre de mémoire  $p_n$  dépendant du temps. L'approche la plus simple pour essayer de traiter ce cas semble être les martingales. En effet, on conserve l'équation

$$\forall n \geq 1, \quad \mathbb{E}(S_{n+1} | \mathcal{F}_n) = \gamma_n S_n,$$

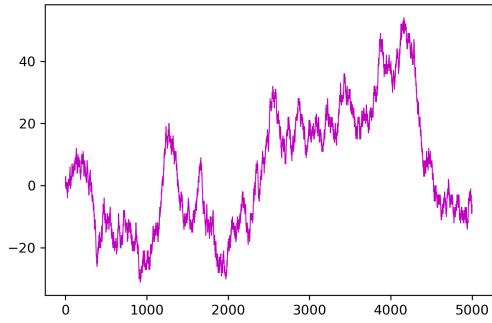
où l'on doit simplement changer  $\gamma_n$ , qui vaut désormais  $\gamma_n = 1 + \frac{2p_n - 1}{n}$ . On peut donc de nouveau créer une martingale à partir de  $(S_n)_{n \geq 0}$ , en posant pour  $n \geq 0$ ,  $M_n = a_n S_n$ , avec  $a_0 = a_1 = 0$ , et pour  $n \geq 2$ ,  $a_n = \prod_{k=1}^{n-1} \gamma_k^{-1}$ .

Cependant la série  $\sum_k a_k^2$  dépend de la suite  $(p_n)_{n \geq 1}$  et devient donc beaucoup plus compliquée à étudier. Nous allons donc nous contenter d'observations quelques simulations de marches pour différentes suites  $(p_n)_{n \geq 1}$ .

Tout d'abord, observons des simulations de marches après 5000 pas pour les suites  $(\frac{1}{n})_{n \geq 1}$  et  $(\frac{1}{\sqrt{n}})_{n \geq 1}$ .



(a) Marche pour la suite  $(\frac{1}{n})_{n \geq 1}$



(b) Marche pour la suite  $(\frac{1}{\sqrt{n}})_{n \geq 1}$

On voit que ces trajectoires ressemblent plutôt au régime diffusif étudié précédemment, ce qui semble plutôt cohérent puisque pour les deux suites, pour tout  $n \geq 2$ ,  $p_n \leq \frac{3}{4}$ . Essayons donc d'utiliser une suite qui reste plus longtemps au dessus du seuil des différents régimes étudiés précédemment. Nous allons pour cela utiliser la suite  $(\frac{\log(n)}{n^{\frac{1}{3}}})_{n \geq 1}$ . Parmi les premiers termes, certains termes dépassent 1, nous allons donc prendre pour  $n \geq 1$ , le minimum entre 1 et  $\frac{\log(n)}{n^{\frac{1}{3}}}$ .

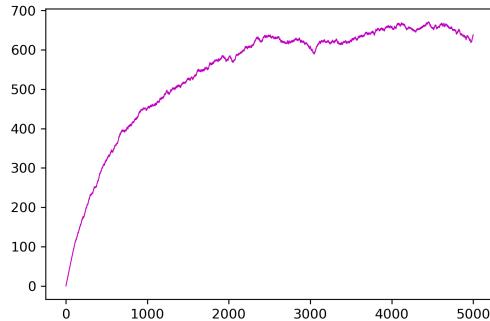


FIGURE 15 – Marche pour la suite  $(\min(1, \frac{\log(n)}{n^{\frac{1}{3}}}))_{n \geq 1}$

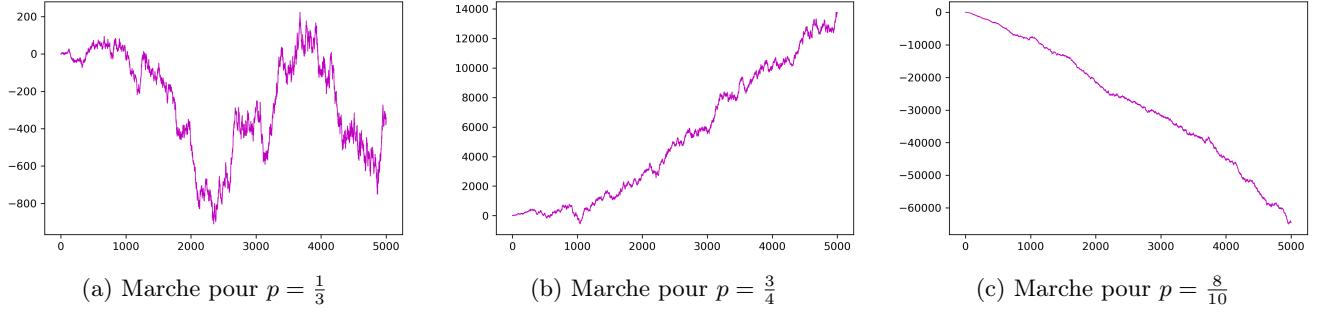
On voit que cette fois-ci la trajectoire ressemble plus à celles des régimes critique et super-diffusif étudiés précédemment, ce qui semble de nouveau cohérent puisque la valeur de  $p_n$  reste longtemps au dessus de  $\frac{3}{4}$ , et la suite tend ensuite lentement vers zéro. On peut donc imaginer que l'on pourrait obtenir des critères sur la suite  $(p_n)_{n \geq 1}$  afin de déterminer le comportement asymptotique de la suite.

Comme autre piste de généralisations, on pourrait envisager qu'au lieu de ne refaire qu'un pas parmi les précédents, l'éléphant refait une séquence du passé. Pour formaliser cela, on peut se donner un  $M \in \mathbb{N}^*$ , qui fixe la taille des séquences dont on se souvient. Ainsi, les  $M + 1$  premiers pas sont faits au hasard, tous selon une loi de Rademacher d'un certain paramètre  $q \in [0, 1]$ , puis pour toute étape  $n \geq M + 2$ , l'éléphant se souvient uniformément au hasard d'un moment parmi les étapes  $[M + 1, n - 1]$ , et refait la séquence  $X_{k-M}, \dots, X_k$  avec probabilité  $p$ , et  $-X_{k-M}, \dots, -X_k$  avec probabilité  $1 - p$ . Cela signifie que le  $n$ -ième vaut  $S_k - S_{k-M}$  avec probabilité  $p$ , et  $S_{k-M} - S_k$  avec probabilité  $1 - p$ . Les approches utilisant les urnes et les arbres ne semblent plus utilisables,

l'approche martingale semble donc être la plus envisageable mais par manque de temps elle n'a pas été développée. Nous allons donc nous contenter d'observer de nouveau des simulations de trajectoires.

Nous allons nous limiter pour les simulations à  $M = 2$  pour ne pas alourdir notre propos avec trop de graphiques, mais les simulations pour des  $M$  supérieurs semblent donner le même type de résultats.

Observons tout d'abord des simulations pour des paramètres de mémoires correspondant aux trois régimes de la marche classique.



On voit que les régimes observés pour la marche classique ont l'air de coller avec ce que l'on observe. Cependant, le basculement entre régimes diffusif et super-diffusif semble en réalité ne plus être pour  $p = \frac{3}{4}$ . En effet, en simulant une trajectoire pour des  $p$  légèrement inférieurs, on voit que l'on obtient déjà des trajectoires ressemblant à un régime super-diffusif.

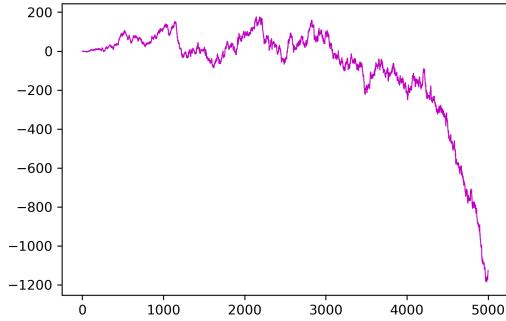


FIGURE 17 – Marche pour  $p = \frac{2}{3}$

Ainsi, on peut envisager que le comportement asymptotique de cette marche se divise en plusieurs régimes, mais que le seuil entre les régimes dépend cette fois ci du paramètre  $M$ .

Enfin, une autre piste envisageable pour modifier la marche de l'éléphant est d'introduire des biais directionnels. Cela peut tout d'abord se traduire en dimension 1 toujours, en imaginant que le paramètre de mémoire n'est pas le même si l'éléphant se souvient d'un pas à droite ou d'un pas à gauche. On aurait donc deux paramètres de mémoire :  $p_d$  et  $p_g$ . L'approche martingale peut-être une bonne option pour étudier ce cas, puisque les arbres ne semblent pas adéquat, et l'utilisation des urnes de Polya est rendue plus difficile puisque la matrice de remplacement devrait dépendre de ce que l'on vient de tirer. De nouveau nous allons observer quelques simulations afin de se faire une idée de ce qu'il pourrait se passer avec ce nouveau modèle. Voici donc plusieurs trajectoires de marches après 5000 pas, pour différentes valeurs de  $p_g$  et  $p_d$ .

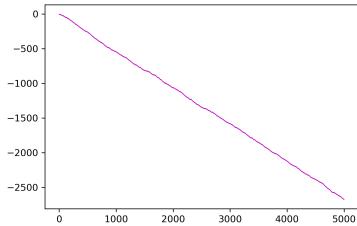


FIGURE 18 –  $p_g = \frac{8}{10}$  et  $p_d = \frac{1}{3}$

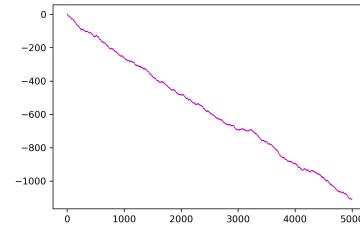


FIGURE 19 –  $p_g = \frac{3}{4}$  et  $p_d = \frac{5}{8}$

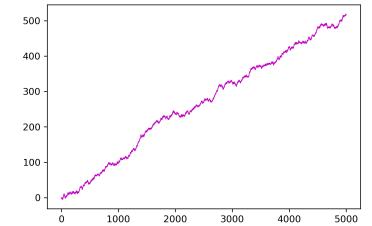


FIGURE 20 –  $p_g = \frac{1}{6}$  et  $p_d = \frac{1}{3}$

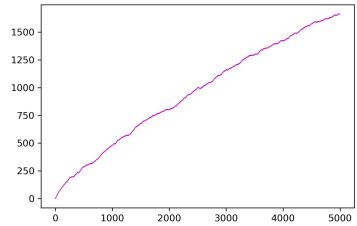


FIGURE 21 –  $p_g = \frac{28}{30}$  et  $p_d = \frac{27}{30}$

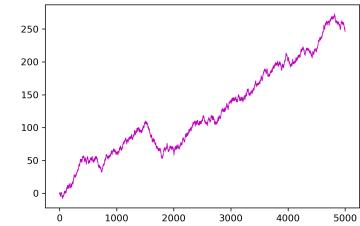


FIGURE 22 –  $p_g = \frac{1}{12}$  et  $p_d = \frac{2}{12}$

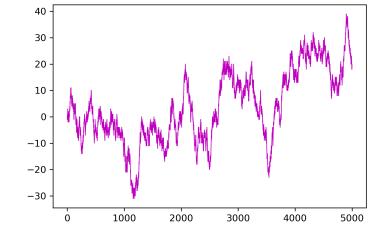


FIGURE 23 –  $p_g = \frac{1}{30}$  et  $p_d = \frac{2}{30}$

On observe différentes choses grâce à ces simulations. On voit tout d'abord sur la figure 18 que s'il y a une grande différence entre les deux paramètres, la marche part très vite dans la direction du paramètre le plus grand, et on observe un comportement proche du régime super-diffusif. On observe le même comportement si l'on garde un écart tout de même significatif même si plus faible que précédemment entre les deux paramètres, peu importe que les paramètres soient petits ou non, comme dans les figures 19 et 20 mais également si les deux paramètres ont un très faible écart, mais qu'ils sont tous les deux plutôt grands, comme dans la figure 21. Mais l'on voit que si l'on prend des paramètres à la fois très petits et très proches, comme dans les figures 22 et 23, le comportement s'approche plus du régime critique voire du régime diffusif.

Ainsi, on peut supposer que le comportement asymptotique de cette marche dépend à la fois de la position des deux paramètres mais aussi de l'écart entre ces derniers.

Pour étudier cette histoire de biais, on peut également imaginer qu'en dimension  $d$ , lorsque l'éléphant ne fait pas le pas de son souvenir, il ne choisit pas uniformément au hasard parmi les autres directions. Par exemple en dimension 2, on pourrait imaginer que l'éléphant fait le pas de souvenir avec probabilité  $p$ , un pas dans la direction opposée avec probabilité  $p'$ , et un pas dans l'une des deux autres directions avec la même probabilité  $\frac{1-p-p'}{2}$  pour chacune. Cette extension pourrait s'étudier avec les urnes de Polya : la matrice de remplacement serait l'identité

avec probabilité  $p$  toujours, la matrice  $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$  avec probabilité  $p'$ , ou l'une des deux matrices  $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

et  $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$  avec probabilité  $\frac{1-p-p'}{2}$  pour chacune. On pourrait alors regarder l'espérance de la matrice de remplacement, ses valeurs propres, etc. Nous allons de nouveau nous contenter d'observer quelques simulations.

Voici donc quelques simulations de trajectoires après 5000 pour différents couples  $(p, p')$ .

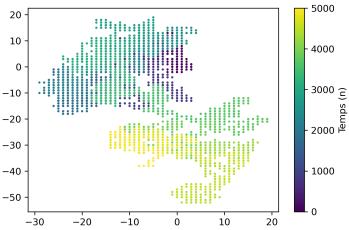


FIGURE 24 –  $p = \frac{3}{4}$  et  $p' = \frac{1}{6}$

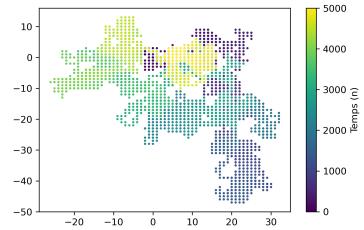


FIGURE 25 –  $p = \frac{3}{4}$  et  $p' = \frac{1}{30}$

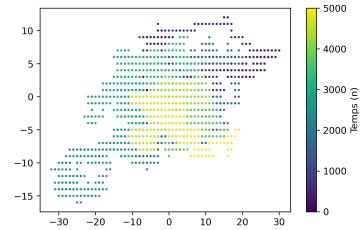


FIGURE 26 –  $p = \frac{8}{10}$  et  $p' = \frac{1}{6}$

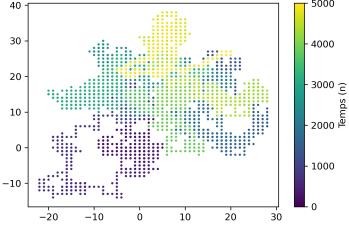


FIGURE 27 –  $p = \frac{8}{10}$  et  $p' = \frac{1}{30}$

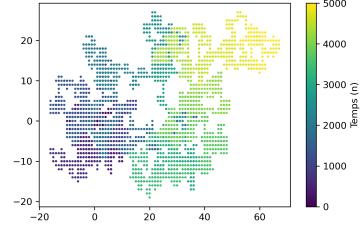


FIGURE 28 –  $p = \frac{1}{3}$  et  $p' = \frac{1}{3}$

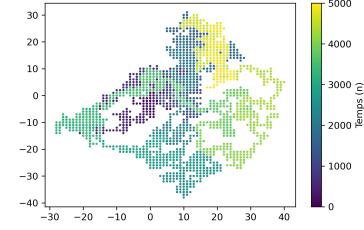


FIGURE 29 –  $p = \frac{1}{3}$  et  $p' = \frac{1}{9}$

Les différentes simulations nous laissent penser que peu importe le couple choisi, le comportement asymptotique de la marche est peu diffusif. En effet, peu importe la position de  $p$ , et peu importe si  $p'$  est au dessus ou en dessous de  $\frac{1-p}{3}$ , c'est à dire que peu importe si l'on attribue plus ou moins de poids à la direction opposée qu'aux deux autres, la marche semble peu s'étendre dans tous les cas.

Il pourrait évidemment y avoir une multitude d'autres pistes afin d'étudier des variations de la marche aléatoire de l'éléphant. Les trois méthodes développées précédemment ne sont pas nécessairement applicables pour toute extension, mais comme nous l'avons vu, elles permettent déjà de traiter beaucoup de cas.

## Références

- Erich Baur and Jean Bertoin. The fragmentation process of an infinite recursive tree and Ornstein-Uhlenbeck type processes. 2015.
- Erich Baur and Jean Bertoin. Elephant random walks and their connection to pólya-type urns. *Physical review E*, 94(5) :052134, 2016.
- Bernard Bercu. A martingale approach for the elephant random walk. *Journal of Physics A : Mathematical and Theoretical*, 51(1) :015201, nov 2017. doi : 10.1088/1751-8121/aa95a6.
- Bernard Bercu and Lucile Laulin. On the multi-dimensional elephant random walk. *Journal of Statistical Physics*, 175 :1146–1163, 2019.
- Marco Bertenghi. Functional limit theorems for the multi-dimensional elephant random walk. *Stochastic Models*, 38(1) :37–50, 2022.
- BM Brown. A general three-series theorem. *Proceedings of the American Mathematical Society*, 28(2) :573–577, 1971.
- Silvia Businger. The Shark Random Swim (Lévy Flight with Memory). *Journal of Statistical Physics*, 172 :701–717, 2018.

- Cristian F. Coletti, Renato Gava, and Gunter M. Schütz. Central limit theorem and related results for the elephant random walk. *Journal of Mathematical Physics*, 58(5) :053303, may 2017. doi : 10.1063/1.4983566.
- Marie Duflo. Random iterative models, stochastic modelling and applied probability, vol. 34, 1997.
- C.C. Heyde. On central limit and iterated logarithm supplements to the martingale convergence theorem. *Journal of Applied Probability*, 14(4) :758–775, 1977.
- Svante Janson. Functional limit theorems for multitype branching processes and generalized pólya urns. *Stochastic Processes and their Applications*, 110(2) :177–245, 2004.
- Naoki Kubota and Masato Takei. Gaussian fluctuation for superdiffusive elephant random walks. *Journal of Statistical Physics*, 177(6) :1157–1171, 2019.
- Rüdiger Kürsten. Random recursive trees and the elephant random walk. *Physical Review E*, 93(3) :032111, 2016.
- Lucile Laulin. Introducing smooth amnesia to the memory of the elephant random walk. *Electronic Communications in Probability*, 27 :1–12, 2022a.
- Lucile Laulin. New insights on the reinforced elephant random walk using a martingale approach. *Journal of Statistical Physics*, 186 :1–23, 2022b.
- Lucile Laulin. *Autour de la marche aléatoire de l’éléphant*. PhD thesis, Bordeaux, 2022c.