

Université Paris 6

Habilitation à diriger des recherches

Spécialité : **Mathématiques**

présentée par M. Philippe Marchal

Théorie des fluctuations, concentration de la mesure, probabilités sur les arbres

soutenue le 22 novembre 2010 devant le jury composé de :

Jean Bertoin (Université Paris 6)
Ron Doney (Université de Manchester) *rapporteur*
Jean-François Le Gall (Université d'Orsay)
Yves Le Jan (Université d'Orsay)
Alain Rouault (Université de Versailles) *rapporteur*
Wendelin Werner (Université d'Orsay)

Remerciements

Tout d'abord, je tiens à remercier les rapporteurs de cette habilitation ainsi que les membres du jury, l'intersection étant non vide.

Seul des trois rapporteurs à ne pas faire partie du jury, Jan Rosinski a permis, par sa diligence et son application dans la rédaction de son rapport, à cette soutenance d'avoir lieu dans des délais raisonnables. Qu'il en soit vivement remercié.

La présence des deux autres rapporteurs, Alain Rouault et Ron Doney, me fait grand plaisir, ce dernier, avec qui j'eus l'occasion de collaborer, se déplaçant spécialement pour l'occasion.

Jean Bertoin et Jean-François Le Gall ont joué un grand rôle dans ma formation de probabiliste. J'eus le second comme directeur de DEA et le premier comme directeur de thèse, vers la fin du XXème mais leurs travaux ultérieurs, notamment sur les processus de fragmentation, coalescents et arbres aléatoires, ont continué de grandement inspirer mes travaux.

Wendelin Werner rédigea un des rapports de ma thèse alors qu'il n'était pas encore autant couvert d'honneurs qu'à présent, et Yves Le Jan fut l'un des examinateurs. J'apprécie particulièrement de les retrouver au jury de cette habilitation.

Bien que la vie normale d'un mathématicien soit celle d'une solitaire misanthrope, je dois beaucoup à la compagnie des collègues que j'ai eus successivement à Paris 6, à l'ENS et maintenant à Paris 13. Qu'ils ne m'en veuillent pas si je ne cite pas leurs noms, la liste étant assez longue. Je me contenterai de mentionner mes coauteurs Patricia Reynaud et Christian Houdré.

Les remerciements d'une habilitation se prêtent naturellement à plus de sobriété que ceux d'une thèse. Je ne saurais toutefois terminer sans évoquer une personne qui était présente à ma soutenance de thèse et qui ne sera pas parmi nous cette fois-ci. C'est à mon frère que je dédie cette habilitation, frère tardivement arrivé et trop tôt disparu – ce fut sa décision ultime.

“ ... und wenn ich einmal ein alter Mann in grauen Haaren bin, dann stehst gerade du mir vielleicht wieder näher als alle Mitlebenden.”

Table des matières

1	Introduction	4
2	Synthèse des travaux de recherche	7
2.1	Concentration de la mesure	7
2.2	Problèmes d'aspect combinatoire	12
2.2.1	Probabilités sur les arbres	12
2.2.2	Marches à boucles effacées	16
2.3	Processus de Lévy, théorie des fluctuations	17
2.3.1	Théorie des fluctuations	17
2.3.2	Processus de Lévy	18
3	Almost sure convergence of the simple random walk to Brownian motion	26
3.1	Introduction	26
3.2	Description of the algorithms	28
3.2.1	The case of the excursion	28
3.2.2	The bridge and the meander	29
3.2.3	The general case	30
3.3	The combinatorial part of Theorem 1	32
3.4	Proof of almost sure convergence	34
3.4.1	Moving steps and local time	34
3.4.2	Martingale properties	35
3.4.3	L^2 estimates	36
3.4.4	The discrete skeleton	36
3.4.5	B is the graph of a continuous function	38
3.4.6	Uniform convergence to B	39
3.5	Further results	40
3.5.1	Trees	40
3.5.2	Urns	40
3.5.3	Random partitions	41
3.5.4	Proof of Theorem 3	42
4	Erratum for the paper "Nested regenerative sets and their associated fragmentation process"	46

Chapitre 1

Introduction

La partie “ Synthèse des travaux de recherche ” s’appuie sur douze articles publiés et un soumis pour publication. Elle présente de manière condensée les résultats obtenus depuis octobre 1999, date de soutenance de la thèse de doctorat. Cette dernière comportait six articles [5, 6, 7, 10, 11, 12]. La bibliographie qui suit donne la liste de tous ces travaux.

En annexe sont donnés deux appendices. Tout d’abord, nous donnons une version étendue de l’article “Constructing a sequence of random walks strongly converging to Brownian motion” [14], la seule différence avec l’original étant la preuve détaillée de la convergence uniforme vers le mouvement brownien, chapitre 3.4.4 à 3.4.6. Puis, dans un erratum pour l’article “Nested regenerative sets and their associated fragmentation process” [15], nous corrigeons une erreur dans le calcul de coefficients relatifs à la fragmentation des cascades de Ruelle.

Bibliographie

- [1] Doney, R. A. ; Marchal, P. A third arc-sine theorem. *Bull. London Math. Soc.* **35** (2003), no. 4, 536–540.
- [2] Houdré C., Marchal P. On the Concentration of Measure Phenomenon for Stable and Related Random Vectors. *Ann. Probab.* **32** (2004), 1496–1508.
- [3] Houdré, Christian ; Marchal, Philippe. Median, concentration and fluctuations for Lévy processes. *Stochastic Process. Appl.* **118** (2008), no. 5, 852–863.
- [4] Houdré, C., Marchal, P, Reynaud–Bouret, P. (2008) Concentration for Norms of Infinitely Divisible Vectors With Independent Components. *Bernoulli* **14** 926–948.
- [5] Marchal, Philippe. Temps d’occupation de $(0, \infty)$ pour les marches aléatoires. *Stochastics Stochastics Rep.* **64** (1998), no. 3-4, 267–282.
- [6] Marchal, Philippe. The best bounds in a theorem of Russell Lyons. *Electron. Comm. Probab.* **3** (1998), 91–94.
- [7] Marchal, Philippe. Distribution of the occupation time for a Lévy process at passage times at 0. *Stochastic Process. Appl.* **74** (1998), no. 1, 123–131.
- [8] Marchal, Philippe. Cycles hamiltoniens aléatoires et mesures d’occupation invariantes par une action de groupe. *C. R. Acad. Sci. Paris S ? ?r. I Math.* **329** (1999), no. 10, 883–886.
- [9] Marchal, Philippe. Loop-erased random walks, spanning trees and Hamiltonian cycles. *Electron. Comm. Probab.* **5** (2000), 39–50.
- [10] Marchal, Philippe. Stable processes on the boundary of a regular tree. *Ann. Probab.* **29** (2001), no. 4, 1591–1611.
- [11] Marchal, Philippe. A combinatorial approach to the two-sided exit problem for left-continuous random walks. *Combin. Probab. Comput.* **10** (2001), no. 3, 251–266.
- [12] Marchal, Philippe. On a new Wiener-Hopf factorization by Alili and Doney. *Séminaire de Probabilités, XXXV*, 416–420, Lecture Notes in Math., 1755, Springer, Berlin, 2001.
- [13] Marchal, Philippe. Two consequences of a path transform. *Bull. London Math. Soc.* **33** (2001), no. 2, 213–220.

- [14] Marchal, Philippe. Constructing a sequence of random walks strongly converging to Brownian motion. *Discrete Math. Theor. Comput. Sci.* 181–190 (2003)
- [15] Marchal, Philippe. Nested regenerative sets and their associated fragmentation process. *Mathematics and computer science*. **III**, 461–470, Trends Math., Birkhäuser, Basel, 2004.
- [16] Marchal, Philippe. Measure concentration for stable laws with index close to 2. *Electron. Comm. Probab.* **10** (2005), 29–35
- [17] Marchal, Philippe. A note on the fragmentation of a stable tree. Fifth Colloquium on Mathematics and Computer Science, **489–499**, *Discrete Math. Theor. Comput. Sci. Proc.* (2008)
- [18] Marchal, Philippe. Small time expansions for transition probabilities of some Lévy processes. *Electron. Commun. Probab.* **14** (2009), 132–142.

Chapitre 2

Synthèse des travaux de recherche

Nous présentons brièvement l'essentiel des résultats obtenus postérieurement à la thèse de doctorat soutenue en octobre 1999, tout en incluant dans la bibliographie les articles qui y figuraient [39, 40, 41, 44, 45, 46]. Les énoncés ont souvent été simplifiés dans un but de clarté. Pour plus de détails, on se reportera aux versions publiées, chacun des travaux dont nous parlons ayant fait l'objet d'un article paru.

2.1 Concentration de la mesure

Le phénomène de *concentration de la mesure* a fait l'objet de recherches intensives chez les probabilistes au cours des vingt dernières années. Il est naturellement impossible de donner un panorama ne serait-ce que succinct des avancées récentes dans ce domaine. On pourra se reporter aux références citées dans le livre de Ledoux [36]. À la différence des résultats asymptotiques où l'on étudie une certaine probabilité P_t indexée par un paramètre t et où l'on obtient des résultats de la forme $P_t \sim f(t)$ pour une certaine fonction $f(t)$, les résultats de concentration de la mesure cherchent à donner des bornes explicites à t fixé. Un exemple typique est le cas des lois stables, où l'on a des estimations asymptotiques précises [12, 68] mais où des majorations non asymptotiques peuvent se révéler utiles (voir par exemple Ben Arous-Guionnet [6] dans le cadre des matrices aléatoires, ou encore [27]).

Nos résultats ont été obtenus en grande partie en collaboration avec Christian Houdré. Ils s'appuient sur une formule de covariance qui s'applique pour les lois indéfiniment divisibles [31, 26]. Ce cadre, assez particulier, englobe cependant les cas gaussiens et exponentiels qui avaient fait l'objet de travaux fondateurs de Borell, Sudakov-Tsirelson et Talagrand [14, 64, 65].

Dans le premier article [26] où il faisait usage de cette formule, Houdré avait établi des résultats de concentration pour les lois indéfiniment divisibles ayant

des moments exponentiels. Dans ce cadre, on sait que les meilleures bornes qu'on puisse espérer sont exponentielles si la mesure de Lévy décroît exponentiellement ; si la mesure de Lévy a un support borné, ce qui est par exemple le cas d'une loi de Poisson, la "bonne" majoration est en $\exp(-x \log x)$. La méthode de covariance permet d'obtenir ces bornes optimales.

Nous avons obtenu des résultats pour les lois stables [28] qui peuvent s'exprimer comme suit. Tout d'abord, il nous faut normaliser la variable aléatoire, ce qu'on peut faire fixant la variance tronquée :

$$\int_{|x|<1} x^2 \Pi(dx) = 1, \quad (2.1)$$

où Π est la mesure de Lévy. On a alors :

Théorème 1 (Houdré-M) *Soit $X \in \mathbb{R}$ une variable aléatoire stable normalisée par (2.1), d'indice $\alpha \in (0, 2)$. Soient f une fonction 1-lipschitzienne et $m(f)$ sa médiane. Alors pour tout $x > 0$,*

$$\mathbb{P}(f(X) - m(f(X)) \geq x) \leq \frac{32e}{\alpha x^\alpha}.$$

La preuve utilise une méthode de troncation. Si nous envisageons la variable stable X comme la loi d'un processus stable (Y_t) au temps 1, nous pouvons tronquer les sauts d'une taille supérieure à un certain seuil choisi de manière adaptée. Cela donne une décomposition de la forme $Y = Y^{(1)} + Y^{(2)}$ où $Y^{(1)}$ a des petits sauts et $Y^{(2)}$ des grands sauts. La formule de covariance s'applique à $Y^{(1)}$ comme dans [26] et par ailleurs, on peut majorer de manière expéditive la probabilité $\mathbb{P}(Y^{(2)} = 0)$.

Dans le cas où α est proche de 2, du fait de la normalisation choisie, la loi stable est "proche" d'une gaussienne pour laquelle on a une queue en $\exp(-x^2/2)$ alors que la queue de la loi stable est en $x^{-\alpha}$ pour $\alpha < 2$. Il y a donc là une sorte de "transition de phase" que nous analysons un peu plus précisément dans [50]. Nous montrons que pour α proche de 2, la borne gaussienne est valable dans une fenêtre de taille $\sqrt{\log(1/(2-\alpha))}$.

Théorème 2 (i) *Avec les mêmes hypothèses que dans le théorème 1, si de plus $\alpha > 3/2$, posant $M = 1/(2-\alpha)$, on a*

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \geq x) \leq \frac{(1+8e^2)}{Mx^\alpha}$$

dès que

$$x^\alpha \geq 4 \log M \log(1+2M \log M).$$

(ii) *Soient $a < 1$, $a', \varepsilon > 0$. Alors si α est assez proche de 2,*

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \geq x) \leq (1+\varepsilon)e^{-x^\alpha/2}$$

dès que

$$a' \leq x^\alpha \leq a \log M.$$

Une autre forme de transition de phase peut être observée pour des processus asymptotiquement stables [29]. Ces processus sont ceux sans partie brownienne, dont la mesure de Lévy, au voisinage de 0, “ressemble” à celle d’une loi stable d’indice $\alpha < 2$ mais dont la variance est bornée. En temps petit, ces processus se comportent comme des processus stables d’indice α cependant qu’en temps grand, il se comportent comme un mouvement brownien. La transition entre ces deux types de comportement peut être analysée de manière plus détaillée à l’aide des quantités suivantes, définies pour $R > 0$:

$$\begin{aligned} V^2(R) &= \int_{\|x\| \leq R} \|x\|^2 \Pi(dx), \\ M(R) &= \int_{\|x\| > R} \|x\| \Pi(dx), \\ \bar{\Pi}(R) &= \int_{\|x\| > R} \Pi(dx). \end{aligned}$$

Dans le cas stable, ces quantités sont des fonctions de puissance de R . Dans le cas asymptotiquement stable, elles se comportent approximativement comme ces mêmes fonctions de puissance au voisinage de 0. On montre alors qu’à la fois la médiane et les déviations par rapport à la médiane sont contrôlées par la fonction, indicée par $c > 0$,

$$h_c(t) = \inf \left\{ R > 0 : 0 < \frac{V^2(R)}{R^2} \leq \frac{c}{t} \right\}. \quad (2.2)$$

Là encore, h_c est une fonction de puissance dans le cas stable; dans le cas asymptotiquement stable, elle ressemble à cette même fonction de puissance au voisinage de 0 et à la fonction correspondante pour le mouvement brownien au voisinage de l’infini. Pour simplifier les notations, nous énonçons le théorème dans le cas symétrique mais le cas général peut être traité de manière similaire [29].

Théorème 3 (Houdré-M) *Soit X un processus de Lévy symétrique. Soit f une fonction 1-lipschitzienne, soient $c > 0$, $t > 0$ et soit h_c la fonction donnée par la formule (2.2). Alors pour tout $t > 0$ tel que*

$$t\bar{\Pi}(h_c(t)) \leq 1/4,$$

toute médiane $mf(X_t)$ de $f(X_t)$ vérifie :

$$|mf(X_t) - f(0)| \leq h_c(t) [\sqrt{c} + 3g_c(1/4)]$$

où $g_c(x)$ est la solution en y de l’équation

$$y - (y + c) \log \left(1 + \frac{y}{c} \right) = \log x.$$

Certains résultats apparentés, exprimés plutôt en termes d'espérance, sont donnés dans l'article de Marcus-Rosinski [38]. Pour l'étude des fluctuations, on fait l'hypothèse supplémentaire d'existence d'une constante K telle que pour tout $R > 0$,

$$M(R) \leq K \frac{V^2(R)}{R}. \quad (2.3)$$

Le résultat s'exprime en fonction de la condition suivante, qui dépend d'un réel $A > 0$:

$$\bar{\Pi}(R) \leq A \frac{V^2(R)}{R^2}. \quad (2.4)$$

Les déviations de X_t par rapport à son espérance sont alors contrôlées par :

Théorème 4 (Houdré-M) *Sous les mêmes hypothèses qu'au théorème 3, supposons vérifiée (2.3). Soient f une fonction 1-lipschitzienne et $A, q > 0$. Alors pour tout $t > 0$ tel que $R = h_{q/2A}(t)$ satisfasse (2.4) et pour tout x tel que*

$$x \geq \left[\frac{qK}{2A} + g_{q/2A}(q/2) \right] h_{q/2A}(t),$$

on a :

$$\mathbb{P}(f(X_t) - \mathbb{E}f(X_t) \geq x) \leq q$$

et

$$\mathbb{P}(f(X_t) - \mathbb{E}f(X_t) \leq -x) \leq q.$$

Enfin, nous donnons des résultats de concentration pour des espaces produits. Le cadre général est le suivant : on a une variable aléatoire $X \in \mathbb{R}^n$ à coordonnées indépendantes et iid et on veut démontrer des inégalités de la forme

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq y) \leq G(y)$$

pour certaines fonctions $F : \mathbb{R}^n \rightarrow \mathbb{R}$, où la fonction G ne dépend pas de n . Naturellement, pour qu'une telle inégalité soit vérifiée, il est nécessaire d'imposer certaines conditions. Par exemple, au vu du théorème de la limite centrale, si on étudie la norme ℓ^p de X en supposant les coordonnées iid et l'existence de moments d'ordre $2p$,

$$\begin{aligned} \|X\|_p &= \left(\sum_{i=1}^n |X_i|^p \right)^{1/p} \simeq (n\mathbb{E}|X_1|^p + \sqrt{n}Z)^{1/p} \\ &\simeq (n\mathbb{E}|X_1|^p)^{1/p} + n^{(1/p)-(1/2)} \frac{(\mathbb{E}|X_1|^p)^{1/p}}{p} Z, \end{aligned}$$

où Z est une gaussienne de variance $\mathbb{E}|X_1|^{2p}$, le signe \simeq étant à prendre ici dans un sens assez vague. Les fluctuations sont donc d'ordre $n^{(1/p)-(1/2)}$ et ne peuvent être bornées indépendamment de la dimension que si $p \geq 2$. Il est donc

naturel de supposer que la fonction F est k -Lipschitzienne, k fixé, pour la norme euclidienne.

De nombreux travaux se sont attachés à donner des résultats sous diverses hypothèses pour la loi de X . Ainsi, des inégalités fonctionnelles du type Poincaré, Sobolev logarithmique, transport, isopérimétrie, log-concavité, permettent d'obtenir des résultats dans cette direction, avec une borne exponentielle ou gaussienne suivant les cas. L'article fondateur de Talagrand [65] a suscité une littérature très abondante sur le sujet, nous renvoyons de nouveau à [36] pour des références.

Pour une probabiliste, ces inégalités fonctionnelles ne sont pas forcément très naturelles, bien qu'elles soient vérifiées à chaque fois dans le cas gaussien. Dans un travail avec Christian Houdré et Patricia Reynaud [30], nous avons obtenu des résultats dans le cadre, relativement usuel en probabilités, où X a des coordonnées iid, indéfiniment divisible, et où la fonction F est la norme ℓ^p avec $2 \leq p < \infty$, ou encore la norme euclidienne d'une projection (dans ce dernier cas, on a cependant une légère dépendance en la dimension).

Théorème 5 *Soit X une variable aléatoire dans \mathbb{R}_+^n à coordonnées iid et indéfiniment divisibles. Alors pour tout $p \geq 2$ et tout $s > 0$,*

$$\mathbb{P}(\|X\|_p - \mathbb{E}\|X\|_p \geq s) \leq \exp\left(-2 \int_0^{s/2} h_p^{-1}(t) dt\right)$$

où h_p^{-1} est l'inverse de la fonction h_p donnée par

$$h_p(t) = p^2 \mathbb{E} \left[\left(\left(1 + \frac{2X_1}{m_p^{1/p}} \right)^{2p-2} + 2^{2p} \frac{m_{2p}}{m_p^2} \right) X_1 (e^{tX_1} - 1) \right]$$

et où, pour tout $q > 0$, les quantités m_q s'expriment en fonction des moments de X_1 d'ordre $\leq q$.

Comme nous l'avons mentionné plus haut, dans le cas où la mesure de Lévy est à support borné, la meilleure borne qu'on puisse espérer est en $\exp(-x \log x)$. Une version quantitative plus précise est fournie par le théorème suivant :

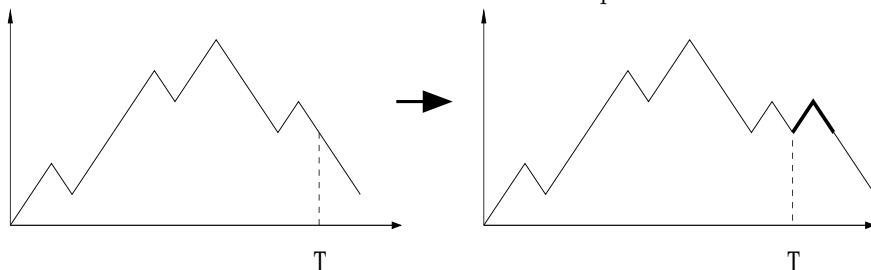
Théorème 6 (Houdré-M-Reynaud) *Soit X une variable aléatoire dans \mathbb{R}^n à coordonnées iid, indéfiniment divisibles, de mesure de Lévy Π et sans partie gaussienne. Supposons que Π soit à support borné et posons $R = \inf\{\rho : \Pi(|x| > \rho) = 0\}$. Alors*

$$\mathbb{E} \left[e^{\frac{\|X\|_2}{R} \log\left(\frac{\lambda \|X\|_2}{R}\right)} \right] < +\infty,$$

pour tout λ tel $\lambda V^2 / R^2 < 1/e$, où $V^2 = 8 \int_{\mathbb{R}} |u|^2 \Pi(du)$.

Ces résultats semblent pouvoir se généraliser dans un cas très général de vecteurs à coordonnées iid, en supposant seulement l'existence de moments exponentiels mais pas l'indéfinie visibilité. En revanche, des tentatives d'introduire des classes plus générales de fonctions F (par exemple les fonctions convexes) n'ont pas abouti mais c'est naturellement une question à creuser.

FIG. 2.1 – On insère un “chapeau”



2.2 Problèmes d’aspect combinatoire

2.2.1 Probabilités sur les arbres

Nous donnons dans [48] une construction presque sûre du mouvement brownien à partir de la marche aléatoire simple, fondée sur un algorithme engendrant récursivement des arbres binaires dû à Rémy [59]. Cette construction est plus explicite que celle de Komlós-Major-Tusnády [33, 34] mais donne des approximations moins fines.

La construction de Rémy est très simple : on part de l’arbre à deux sommets puis, récursivement, on choisit une arête uniformément au hasard et on “greffe” une feuille sur cette arête. Il est facile de voir qu’on obtient au bout de n étapes un arbre binaire aléatoire uniforme à $n + 1$ feuilles.

Il existe une correspondance combinatoire bien connue entre les arbres et les chemins de Lukasiewicz, chemins dont les sauts sont à valeurs entières et supérieurs à -1 . Dans cette correspondance, les arbres binaires deviennent des chemins avec des pas ± 1 , donc des chemins de Dyck pour les combinatoriciens ou encore des trajectoires de la marche aléatoire simple sur \mathbb{Z} pour les probabilistes.

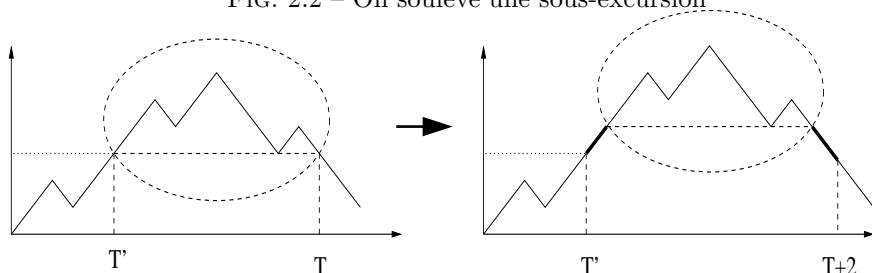
L’algorithme de Rémy devient ainsi, par la correspondance de Lukasiewicz, une construction récursive de trajectoires de marches aléatoires. A priori, on obtient de la sorte des excursion mais des généralisations immédiates permettent de construire des ponts, (des autoroutes) ou des trajectoires non conditionnées.

Notons que cette construction ajoute, à chaque étape, *deux* pas et non un seul. Grosso modo, à chaque étape, on choisit un temps T au hasard et soit on y ajoute un “chapeau”, soit on y soulève une sous-excursion. Voir les figures 1 et 2.

On montre, par des arguments de martingales, qu’il y a convergence presque sûre, la limite ne pouvant bien sûr être qu’un mouvement brownien. Nous donnons en appendice une version étendue de l’article paru originellement [48]. Cette version donne une démonstration plus détaillée de la convergence uniforme :

Théorème 7 *Presque sûrement, la suite de trajectoires de marche aléatoire obtenue par l’algorithme de Rémy converge uniformément vers une fonction continue $(B_t, 0 \leq t \leq 1)$ dont la loi est celle du mouvement brownien.*

FIG. 2.2 – On soulève une sous-excursion



On pourrait obtenir des résultats analogues pour les versions conditionnées (pont ou excursion). Il se trouve que cette convergence presque sûre a lieu pour les arbres eux-mêmes, la limite est l'arbre continu d'Aldous [1, 2, 3].

Nous donnons dans [51] une généralisation de l'algorithme de Rémy qui permet de construire les arbres stables étudiés par Duquesne-Le Gall [20]. Notons qu'une généralisation supplémentaire a été introduite récemment par Chen-Ford-Winkel [16].

On fixe un paramètre $\alpha \in [1, 2]$, on part toujours de l'arbre à deux sommets puis, récursivement, on choisit au hasard avec probabilité proportionnelle au poids :

- soit une arête, avec poids $\alpha - 1$,
- soit un sommet de degré d , avec poids $d - 1 - \alpha$.

Puis on “greffe” une feuille à l'endroit choisi. Le cas $\alpha = 2$ correspond à l'algorithme de Rémy.

On obtient au bout de n étapes un arbre à $n + 1$ feuilles dont la loi est exactement celle du squelette d'un arbre stable, qu'on obtient en échantillonnant uniformément $n + 1$ points de la frontière. La loi de ce squelette avait été donnée dans [20]. Par ailleurs, cet arbre peut être vu comme un arbre de Galton-Watson conditionné par le nombre total de feuilles.

On montre, toujours par des arguments de martingale, une convergence au sens faible :

Proposition 1 *Pour tout couple de sommets (u, v) dans la suite d'arbres construits par la généralisation de l'algorithme de Rémy, notons $d(u, v)$ la distance de graphe. Alors la distance renormalisée*

$$\frac{d(u, v)}{n^{(\alpha-1)/\alpha}}$$

converge presque sûrement vers une variable aléatoire non dégénérée.

On obtiendrait de même la convergence des distances pour tout k -uplet de points. Nous n'avons pas établi, en revanche, le caractère uniforme de cette convergence, ce qui donnerait la convergence dans la topologie de Gromov-Hausdorff, voir par exemple Evans [21]. Des travaux récents de Haas *et al.* étudient la convergence pour cette topologie dans le cadre plus général des

arbres de fragmentation [24, 25] mais ne donnent cependant pas la convergence presque sûre au sens de Gromov-Hausdorff.

Les arbres stables, comme précisé plus haut, peuvent être vus comme des arbres de fragmentation, au sens de la théorie générale des fragmentations introduite par Bertoin [9]. On peut en effet considérer les feuilles d'un arbre fini (ou les branches infinies dans le cas d'un arbre infini) comme portant une certaine masse. Ainsi chaque noeud interne possède une masse en-dessous de lui, masse qui se divise en plusieurs fragments associés aux sous-arbres issus de ce noeud interne. La propriété de branchement des squelettes des arbres stables implique que les dislocations associées aux différents noeuds internes sont indépendantes, induisant un processus de fragmentation.

Notre construction permet d'analyser les dislocations de ce processus à l'aide du modèle du restaurant chinois de Pitman [56]. Ce modèle fournit une approche discrète des lois de Poisson-Dirichlet, qui apparaissent par ailleurs dans le cadre des processus continus (voir entre autres Pitman-Yor [57]).

Dans ce modèle, normalement paramétré par deux réels $\alpha \in [0, 1]$ et $\beta \geq -\alpha$, des clients arrivent un par un dans un restaurant et s'installent, avec probabilité proportionnelle au poids :

- soit à une table déjà occupée, ceci avec poids $q - \alpha$ où q est le nombre de clients déjà assis à cette table,
- soit à une nouvelle table, ceci avec poids $r\alpha + \beta$ où r est le nombre de tables déjà occupées.

On montre que quand le nombre de clients tend vers l'infini, la proportion de clients à chaque table converge presque sûrement, engendrant ainsi une famille aléatoire de réels positifs de somme 1, ce qu'on appelle une partition de masse.

Nous utilisons une variante de ce modèle avec un paramètre de plus. On notera

$$R(\alpha, \beta, n)$$

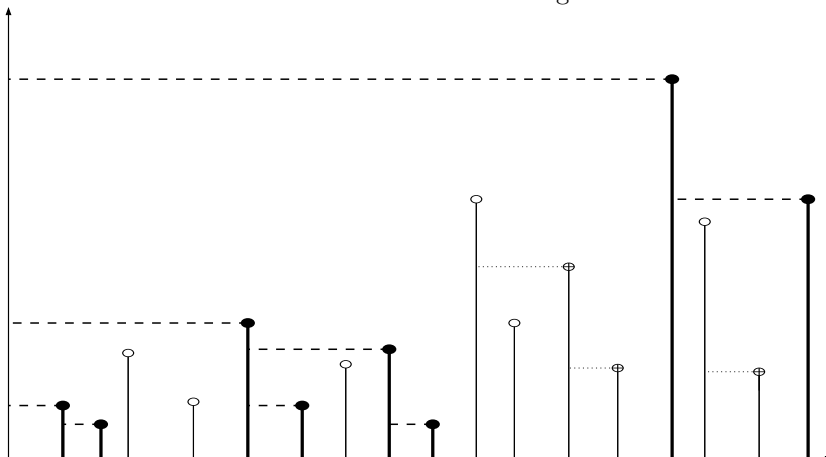
la mesure de probabilités sur les partitions de masse obtenues comme fréquences asymptotiques des tables d'un restaurant chinois, dans le cas où il y a initialement n clients à la première table et 1 client à la deuxième table, processus bien défini dès que $\beta \geq -2\alpha$.

La construction récursive des arbres stables donne une représentation de la mesure de dislocation comme somme de mesures associées au restaurant chinois, ce qui permet d'expliquer pourquoi cette mesure peut être interprétée comme un restaurant chinois de paramètres $(1/\alpha, -1)$, ce jeu de paramètres n'ayant *a priori* pas de sens. Cette représentation permet de retrouver les résultats obtenus par Miermont [54] qui utilisait la représentation en termes de processus stables totalement asymétriques.

Théorème 8 *La fragmentation associée aux arbres stables a pour mesure de dislocation*

$$\sum_{n \geq 1} p_n(\alpha) R(1/\alpha, -1, n),$$

FIG. 2.3 – L'ensemble régénératif



avec $p_1(\alpha) = 1$ et pour $n \geq 2$,

$$p_n(\alpha) = \prod_{i=1}^{n-1} \left(1 + \frac{\alpha - 1}{i\alpha}\right) - \prod_{i=1}^{n-2} \left(1 + \frac{\alpha - 1}{i\alpha}\right).$$

Une autre situation où nous représentons une mesure de dislocation comme somme de mesures associées au restaurant chinois est celle des cascades de Ruelle [49]. Celles-ci ont été introduites dans le cadre de l'étude des verres de spins [61], voir Bolthausen-Sznitman [13] pour le lien avec le modèle SK et Bovier-Kourkova [15] pour une analyse complète du modèle GREM. La relation avec un processus de branchement introduit par Neveu a été analysée par Bertoin-Le Gall [11].

Nous construisons une famille d'ensembles régénératifs emboîtés, chacun de ces ensembles régénératifs étant stable et l'indice variant de 0 à 1. La construction utilise un processus de Poisson ponctuel sur \mathbb{R}_+^2 et des variables uniformes associées aux points du processus. Pour un paramètre $\alpha \in [0, 1]$ donné, on colore les points du processus en noir si la variable associée est inférieure à α . Puis on dit qu'un point de \mathbb{R}_+ percole si on peut "remonter en escalier" le long de points noirs jusqu'à l'axe des ordonnées. Voir la figure 3.

Proposition 2 *L'adhérence de l'ensemble des points qui percolent au niveau α forme un ensemble régénératif stable d'indice α et quand α varie, ces ensembles sont emboîtés.*

Les complémentaires de ces ensembles régénératifs forment une famille d'intervalles ouverts et les longueurs de ces intervalles engendrent une partition de masse. En faisant varier α , on obtient un processus de fragmentation/coalescence et la construction permet d'analyser ce processus à l'aide du restaurant chinois. On obtient ainsi une représentation de la mesure de dislocation, cette mesure variant avec le temps.

Théorème 9 *Les cascades de Ruelle peuvent être vues comme un processus de fragmentation inhomogène, de mesure de dislocation au temps α donnée par*

$$\mu_\alpha = \frac{1}{1-\alpha} \sum_{n=1}^{\infty} c_n R(\alpha, -\alpha, n),$$

où les c_n sont définis par :

$$\sum_{n=1}^{\infty} c_n t^n = \frac{t}{1 + \log(1-t)}.$$

Notons qu'une autre description de la mesure de dislocation a été donnée par Basdevant [5]. Pour d'autres aspects de cette fragmentation, voir par exemple Goldschmidt-Martin [23].

2.2.2 Marches à boucles effacées

Nous analysons certains aspects des marches à boucles effacées dans deux articles. Rappelons que, pour une chaîne de Markov sur un espace d'états fini, la marche à boucles effacées est la trajectoire qu'on trace en effaçant les boucles au fur et à mesure qu'elles apparaissent. Ces marches jouent un rôle fondamental dans les problèmes d'invariance conforme, voir par exemple Kenyon [32] pour les liens avec les modèles de dimères et Lawler-Schramm-Werner [35] pour les relations avec le SLE₂.

Ces marches permettent de simuler des arbres couvrants aléatoires grâce à l'algorithme de Wilson [58]. Cet algorithme construit une première branche de l'arbre couvrant en traçant la marche à boucles effacées d'un premier point à la racine de l'arbre, puis construit une deuxième branche en traçant la marche à boucles effacées d'un deuxième point jusqu'à la première branche etc. Parmi les applications, citons l'analyse des forêts couvrantes de \mathbb{Z}^d par Benjamini-Lyons-Peres-Schramm [7].

Dans [43], utilisant une analyse en temps de dernier passage, nous donnons des expressions matricielles pour la loi jointe de l'arbre engendré et des temps d'occupation des différents points de l'espace d'états, ce qui redémontre que cette loi ne dépend pas de l'ordre choisi pour exécuter l'algorithme.

Théorème 10 *Soit τ le temps d'exécution de l'algorithme de Wilson, T l'arbre engendré, M la matrice $n \times n$ de transition où on a enlevé la ligne et la colonne correspondant à la racine. Alors pour tout arbre couvrant orienté U ,*

$$\mathbb{E}(t^\tau \mathbf{1}_{T=U}) = \frac{w(U)t^n}{\det(I - tM)},$$

où $w(U)$ est le poids de U , c'est-à-dire le produit des probabilités des arêtes orientées de U .

Dans [42], nous montrons qu'en arrêtant ces marches à boucles effacées au premier instant où elles produisent un cycle hamiltonien, on obtient le même genre de résultat. En effet, nous donnons la loi jointe du cycle hamiltonien et des temps d'occupation des différents points de l'espace d'états, montrant qu'il ne dépend pas du point de départ choisi.

Théorème 11 *Soit τ le temps d'exécution de l'algorithme des cycles hamiltoniens, H le cycle hamiltonien engendré, M la matrice $n \times n$ de transition. Alors pour tout cycle hamiltonien C ,*

$$\mathbb{E}(t^\tau \mathbf{1}_{M=C}) = \frac{w(C)t^n}{\det(I - tM) + \sum_{C'} w(C')t^n},$$

où la somme au dénominateur est sur tous les cycles hamiltoniens possibles.

Ce résultat sur les cycles hamiltoniens généralise un résultat de Pitman [55] sur le mouvement brownien et les marches aléatoires sur le cercle : si on fait partir un mouvement brownien d'un point A du cercle et l'arrêtons dès qu'il revient en A en ayant fait un tour complet, la mesure d'occupation ne dépend pas du point d'origine.

Il faut remarquer que ces résultats sont interprétables à la lumière de la théorie générale des empilements de pièces due à Viennot [66], cette théorie s'appliquant ici pour les empilements de cycles (on empile les cycles effacés). Cette théorie permet également de réinterpréter certains résultats de Le Jan sur les mesures de boucles, la décomposition d'un empilement comme assemblée de pyramides impliquant la représentation poissonnienne donnée dans [37].

2.3 Processus de Lévy, théorie des fluctuations

2.3.1 Théorie des fluctuations

La théorie des fluctuations s'intéresse aux extremas successifs d'une marche aléatoire ou d'un processus de Lévy réel, ainsi qu'à des quantités associées comme le temps passé positif. En dehors des résultats contenus dans la thèse, nous présentons deux travaux [47, 18].

Considérons tout d'abord, pour une marche aléatoire $(S_k, 0 \leq k \leq n)$ à valeurs réelles, L_n le nombre de temps d'échelle, c'est-à-dire le nombre de fois où on bat un record vers le haut avant l'instant n , en comptant l'instant 0. Considérons \hat{L}_n la quantité analogue pour la marche duale \hat{S} définie par $\hat{S}_k = S_{n-k} - S_n$ pour $0 \leq k \leq n$. On a alors :

Théorème 12 *Le produit $L_n \hat{L}_n$ vérifie*

$$\mathbb{E}(L_n \hat{L}_n) = n + 1.$$

Notons que la dualité des subordonateurs associés aux quantités L_n et \hat{L}_n a été étudiée de manière détaillée par Vigon [67].

Pour un processus de Lévy qui visite les points, ou pour une marche aléatoire à valeurs entières, on peut définir, sans problème un pont (trajectoire conditionnée à revenir en 0) ou, pour une trajectoire non conditionnée, une partie pont et une partie méandre (parties respectivement avant et après la dernière visite de 0).

Nous montrons qu'on peut définir un équivalent de cette notion de pont dans le cas général en considérant des trajectoires revenant entre 0 et la première hauteur d'échelle, c'est-à-dire la hauteur du premier maximum. On a ainsi :

Théorème 13 *Soit (S_n) une marche aléatoire à valeurs réelles, G_n le premier temps d'atteinte du maximum absolu entre les instants 0 et n et H_1 la première hauteur d'échelle. Alors, conditionnellement à $\{0 \leq S_n \leq H_1\}$, G_n est uniformément réparti sur $\{0, 1, \dots, n\}$.*

Un résultat apparenté est une généralisation de la loi de l'arcsinus [18]. Pour un processus de Lévy qui visite les points, ou pour une marche aléatoire à valeurs entières, on peut regarder, à un temps donné, les tailles respectives du pont et du méandre. Si la condition de Spitzer [63] est vérifiée :

$$\mathbb{P}(S_n \geq 0) \rightarrow \rho \in (0, 1)$$

(nous donnons ici la forme simplifiée, démontrée par Doney [17], de cette condition), alors ces tailles suivent asymptotiquement la loi de l'arcsinus généralisée de paramètre ρ .

Si le processus de Lévy ne visite pas les points ou si la marche aléatoire n'est pas à valeurs entières, mais que la condition de Spitzer est vérifiée, on peut énoncer un résultat analogue. Il faut alors considérer J_n le dernier instant avant l'instant n où on a franchi 0 et L_n l'instant entre J_n et n où on est le plus proche de 0. Alors la trajectoire entre l'instant 0 et l'instant L_n fait office de pont et on a :

Théorème 14 (Doney-M) *Si la condition de Spitzer est vérifiée avec un certain paramètre $\rho \in (0, 1)$, alors L_n/n tend vers la loi de l'arcsinus généralisée de paramètre ρ , de densité sur $[0, 1]$*

$$\frac{x^{\rho-1}(1-x)^{-\rho}}{\Gamma(\rho)\Gamma(1-\rho)}.$$

Ces résultats s'appuient sur des démonstrations utilisant des transformations trajectoires analogues à celle utilisée dans [46] pour redémontrer un résultat d'Alili-Doney [4].

2.3.2 Processus de Lévy

Un travail plus récent analyse le comportement en temps petit des probabilités de transition d'un processus de Lévy réel [52]. Il apparaît en effet qu'il existe peu de travaux allant au-delà de l'estimation classique du premier ordre

$$\mathbb{P}(X_t > y) \sim t(\bar{\Pi}(y) + \mathbf{d}) \tag{2.5}$$

pour t petit, $\overline{\Pi}$ désignant la queue de la mesure de Lévy et \mathbf{d} le terme de dérive (voir par exemple [8]). Signalons que dans l'article [62], certaines hypothèses supplémentaires sont nécessaires qui réduisent considérablement le degré de généralité du résultat. Remarquons aussi que les résultats presque sûrs en temps petit, comme par exemple ceux démontrés Bertoin-Doney-Maller [10], ne sont d'aucune aide dans ce domaine. En effet, pour le comportement presque sûr en temps petit, on peut tronquer la mesure de Lévy aussi près qu'on veut de 0 sans changer les résultats mais si on procède de la sorte, on perd complètement la formule (2.5).

Dans un récent article, Figueiroa-Houdré [22] montrent la validité d'estimations à des ordres supérieurs mais sous des hypothèses assez fortes. Par ailleurs, ils montrent que les termes de ces développements asymptotiques font intervenir les dérivées successives de la densité de la mesure de Lévy, ce qui n'est pas évident *a priori*.

Nous montrons que des hypothèses fortes sont effectivement nécessaires pour obtenir des développements asymptotiques de tous ordres, en exhibant des contre-exemples explicites où de tels développements ne sont pas valides. Nous analysons par ailleurs le rôle des dérivées successives de la densité, expliquant en quoi l'invariance par changement d'échelle du mouvement brownien fait apparaître les dérivées d'ordre impair mais pas celle d'ordre pair. Nous nous concentrons sur le cas où le processus de Lévy est de la forme

$$X_t = S_t + Y_t, \quad (2.6)$$

où S est stable et Y est un processus de Poisson composé de mesure de Lévy Π , qu'on pourra supposer pour simplifier de masse 1, S et Y étant indépendants. Ce cadre, quoique simplifié, permet d'observer certains mécanismes à l'oeuvre dans ce problème. Dans le cas où S est un mouvement brownien, on a :

Théorème 15 (i) *Supposons que Π a une densité continue f sur $[y - \delta, y) \cup (y, y + \delta]$, pour un certain $\delta > 0$. Supposons que*

$$f_+ := \lim_{x \rightarrow 0^+} f(y+x) \neq f_- := \lim_{x \rightarrow 0^-} f(y+x).$$

Alors quand $t \rightarrow 0$,

$$\mathbb{P}(X_t \geq y) - t \left[\overline{\Pi}(z) - \frac{\Pi(\{z\})}{2} \right] \sim c(f_- - f_+) \lambda t^{3/2}.$$

(ii) *Plus généralement, supposons qu'il existe un entier $n \geq 1$ tel que pour tout $i < 2n$,*

$$f^{(i)}(y+) = f^{(i)}(y-)$$

mais que

$$f^{(2n)}(y+) \neq f^{(2n)}(y-).$$

Alors il existe des constantes $c_k, 1 \leq k \leq 2n+2$ telles que quand $t \rightarrow 0$,

$$\mathbb{P}(X_t \geq y) = \sum_{k=1}^{n+1} c_k t^k + c_{n+2} t^{n+(3/2)} + o(t^{n+(3/2)}).$$

Les c_k s'expriment en fonction des $f^{(2k-1)}(y)$ pour $k \leq n$.

Dans le cas où S est stable d'indice $\alpha < 2$, on peut construire des contre-exemples analogues à ceux du théorème 17 (i) mais les processus stables ont un effet plus régularisant. En effet :

Théorème 16 *Supposons qu'il existe $\beta > \alpha$, $a \in \mathbb{R}$, $b, \delta_0 > 0$ tels que pour tout $|x| < \delta_0$,*

$$|\bar{\Pi}(y+x) - \bar{\Pi}(y) - ax| < bx^\beta.$$

Alors il existe un réel c tel que quand $t \rightarrow 0$,

$$\mathbb{P}(X_t \geq y) = t [\bar{\nu}(y) - \mathbb{P}(S_1 < 0)\Pi(\{y\})] + ct^2 + o(t^2).$$

Notons que notre étude se réduit au cadre unidimensionnel mais que des extensions multidimensionnelles semblent possibles.

Le théorème 17 peut s'interpréter en termes de la régularité au bord des solutions d'une équation intégro-différentielle. Considérons en effet l'équation

$$\frac{\partial u(t, x)}{\partial t} = \lambda u(t, x) + \frac{\sigma^2}{2} \frac{\partial^2 u(t, x)}{\partial x^2} + \int_{\mathbb{R}} (u(t, x+y) - u(t, x))\Pi(dy),$$

où Π est une mesure de masse 1, avec la condition initiale $u(0, x) = f(x)$. Cette équation décrit l'évolution d'une population : $u(t, x)$ est la population à la position x et à l'instant t , cette population se reproduit à taux λ , une particule qui naît saute à sa naissance suivant une loi de saut donnée par Π puis elle suit un mouvement brownien de variance σ^2 .

On vérifie que la solution de cette équation est

$$u(t, x) = e^{\lambda t} \mathbb{E}_x f(X_t),$$

où X_t est de la forme (2.6), la mesure de Lévy de Y étant Π et la variance du mouvement brownien étant σ^2 . En particulier, supposons qu'on ait $f(x) = \mathbf{1}_{x \geq 0}$. Alors

$$u(t, x) = e^{\lambda t} \mathbb{P}_0(X_t \geq x).$$

La fonction u est naturellement C^∞ sur $(0, \infty) \times \mathbb{R}$ mais en 0, sa régularité dépend de la régularité de la mesure Π . Plus précisément, supposons que Π ait une densité g . Si $g^{(i)}$ est continue en x_0 pour tout $i \leq 2n-1$ et que $g^{(2n)}$ est discontinue en x_0 , alors la fonction

$$x \mapsto \frac{\partial^n u(0, x)}{\partial x^n}$$

est discontinue en x_0 . Si en revanche $g^{(i)}$ est continue en x_0 pour tout $i \leq 2n-1$ et que $g^{(2n+1)}$ est discontinue en x_0 , alors la fonction

$$t \mapsto \frac{\partial^n u(t, x_0)}{\partial t^n}$$

est discontinue en 0. Ainsi suivant la parité de la première dérivée de g ayant une discontinuité, on a une irrégularité spatiale ou temporelle de la solution u .

Bibliographie

- [1] Aldous, David. The continuum random tree. I. *Ann. Probab.* **19** (1991), no. 1, 1–28.
- [2] Aldous, David. The continuum random tree. II. An overview. *Stochastic analysis (Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser., **167**, Cambridge Univ. Press, Cambridge, 1991
- [3] Aldous, David. The continuum random tree. III. *Ann. Probab.* **21** (1993), no. 1, 248–289.
- [4] Alili, L. and Doney, R. A.. Wiener–Hopf factorisation revisited and some applications, *Stochastics Stochastics Rep.* **66** (1999) 87–102.
- [5] Basdevant, A.-L. Ruelle’s probability cascades seen as a fragmentation process. *Markov Process. Related Fields* **12** (2006), no. 3, 447–474.
- [6] Ben Arous, Gérard ; Guionnet, Alice . The spectrum of heavy tailed random matrices. *Comm. Math. Phys.* **278** (2008), no. 3, 715–751.
- [7] Benjamini, Itai ; Lyons, Russell ; Peres, Yuval ; Schramm, Oded. Uniform spanning forests. *Ann. Probab.* **29** (2001), no. 1, 1–65.
- [8] Bertoin, J. *Lévy processes*. Cambridge University Press, 1996.
- [9] Bertoin, Jean. *Random fragmentation and coagulation processes*. Cambridge Studies in Advanced Mathematics, **102**. Cambridge University Press, Cambridge, 2006.
- [10] Bertoin, J. ; Doney, R. A. ; Maller, R. A. Passage of Lévy processes across power law boundaries at small times. *Ann. Probab.* **36** (2008), no. 1, 160–197.
- [11] Bertoin, Jean ; Le Gall, Jean-François. The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields* **117** (2000).
- [12] Bingham N.H., Goldie C.M. and Teugels J.L. *Regular Variation*. Cambridge University Press, Cambridge, 1987.
- [13] Bolthausen, E. ; Sznitman, A.-S. On Ruelle’s probability cascades and an abstract cavity method. *Comm. Math. Phys.* **197** (1998), no. 2, 247–276.
- [14] C. Borell. The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** (1975), 207–216.

- [15] Bovier, Anton ; Kurkova, Irina. Much ado about Derrida's GREM. *Spin glasses*, 81–115, Lecture Notes in Math., 1900, Springer, Berlin, 2007.
- [16] Chen, Bo ; Ford, Daniel ; Winkel, Matthias. A new family of Markov branching trees : the alpha-gamma model. *Electron. J. Probab.* **14** (2009), no. 15, 400–430.
- [17] Doney R. A. Spitzer's condition and ladder variables in random walks, *Prob. Theory Related Fields* **101** (1995) 577–580.
- [18] Doney, R. A. ; Marchal, P. A third arc-sine theorem. *Bull. London Math. Soc.* **35** (2003), no. 4, 536–540.
- [19] Donsker, Monroe D. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* (1951). no. 6.
- [20] Duquesne, Thomas ; Le Gall, Jean-François. Random trees, Lévy processes and spatial branching processes. *Astérisque* No. 281 (2002).
- [21] Evans, Steven N. *Probability and real trees*. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. Lecture Notes in Mathematics, 1920. Springer, Berlin, 2008.
- [22] Figueroa-López, José E. ; Houdré, Christian. Small-time expansions for the transition distributions of Lévy processes. *Stochastic Process. Appl.* **119** (2009), no. 11, 3862–3889.
- [23] Goldschmidt, Christina ; Martin, James B. Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **10** (2005), no. 21, 718–745.
- [24] Haas, B. and Miermont, G. Scaling limits of Markov branching trees. *En préparation*.
- [25] Haas, B. ; Miermont, G. ; Pitman, J. ; Winkel, M. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.* **36** (2008), no. 5, 1790–1837.
- [26] Houdré, C. (2002). Remarks on deviation inequalities for functions of infinitely divisible random vectors. *Ann. Probab.* **30**, 1223–1237. **6** (2001), 15–29.
- [27] Houdré, Christian ; Xu, Hua Concentration of the spectral measure for large random matrices with stable entries. *Electron. J. Probab.* **13** (2008), no. 5, 107–134.
- [28] Houdré C., Marchal P. On the Concentration of Measure Phenomenon for Stable and Related Random Vectors. *Ann. Probab.* **32** (2004), 1496–1508.
- [29] Houdré, Christian ; Marchal, Philippe. Median, concentration and fluctuations for Lévy processes. *Stochastic Process. Appl.* **118** (2008), no. 5, 852–863.
- [30] Houdré, C., Marchal, P, Reynaud-Bouret, P. (2008) Concentration for Norms of Infinitely Divisible Vectors With Independent Components. *Bernoulli* **14** 926–948.

- [31] Houdré, C., Pérez-Abreu, V. and Surgailis, D. (1998). Interpolation, correlation identities and inequalities for infinitely divisible variables. *J. Fourier Anal. Appl.* **4**, 651–668.
- [32] Kenyon, Richard. The asymptotic determinant of the discrete Laplacian. *Acta Math.* **185** (2000), no. 2, 239–286.
- [33] Komlós, J.; Major, P.; Tusnády, G. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** (1975), 111–131.
- [34] Komlós, J.; Major, P.; Tusnády, G. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **34** (1976), no. 1, 33–58.
- [35] Lawler, Gregory F.; Schramm, Oded; Werner, Wendelin. Conformal invariance of planar loop-erased random walks and uniform spanning trees. *Ann. Probab.* **32** (2004), no. 1B, 939–995.
- [36] Ledoux, M. (2001) *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, **89**. American Mathematical Society.
- [37] Le Jan, Y. *Cours de St-Flour*. En préparation.
- [38] Marcus M.B., Rosiński J. L^1 -norms of infinitely divisible random vectors and certain stochastic integrals. *Electron. Comm. Probab.* **6** (2001), 15–29.
- [39] Marchal, Philippe. Temps d'occupation de $(0, \infty)$ pour les marches aléatoires. *Stochastics Stochastics Rep.* **64** (1998), no. 3-4, 267–282.
- [40] Marchal, Philippe. The best bounds in a theorem of Russell Lyons. *Electron. Comm. Probab.* **3** (1998), 91–94.
- [41] Marchal, Philippe. Distribution of the occupation time for a Lévy process at passage times at 0. *Stochastic Process. Appl.* **74** (1998), no. 1, 123–131.
- [42] Marchal, Philippe. Cycles hamiltoniens aléatoires et mesures d'occupation invariantes par une action de groupe. *C. R. Acad. Sci. Paris S  r. I Math.* **329** (1999), no. 10, 883–886.
- [43] Marchal, Philippe. Loop-erased random walks, spanning trees and Hamiltonian cycles. *Electron. Comm. Probab.* **5** (2000), 39–50.
- [44] Marchal, Philippe. Stable processes on the boundary of a regular tree. *Ann. Probab.* **29** (2001), no. 4, 1591–1611.
- [45] Marchal, Philippe. A combinatorial approach to the two-sided exit problem for left-continuous random walks. *Combin. Probab. Comput.* **10** (2001), no. 3, 251–266.
- [46] Marchal, Philippe. On a new Wiener-Hopf factorization by Alili and Doney. *S  minaire de Probabilit  s, XXXV*, 416–420, Lecture Notes in Math., 1755, Springer, Berlin, 2001.
- [47] Marchal, Philippe. Two consequences of a path transform. *Bull. London Math. Soc.* **33** (2001), no. 2, 213–220.

- [48] Marchal, Philippe. Constructing a sequence of random walks strongly converging to Brownian motion. *Discrete Math. Theor. Comput. Sci.* 181–190 (2003)
- [49] Marchal, Philippe. Nested regenerative sets and their associated fragmentation process. *Mathematics and computer science*. **III**, 461–470, Trends Math., Birkhäuser, Basel, 2004.
- [50] Marchal, Philippe. Measure concentration for stable laws with index close to 2. *Electron. Comm. Probab.* **10** (2005), 29–35
- [51] Marchal, Philippe. A note on the fragmentation of a stable tree. Fifth Colloquium on Mathematics and Computer Science, **489–499**, *Discrete Math. Theor. Comput. Sci. Proc.* (2008)
- [52] Marchal, Philippe. Small time expansions for transition probabilities of some Lévy processes. *Electron. Commun. Probab.* **14** (2009), 132–142.
- [53] Massart, P. (2007). Concentration inequalities and model selection. Ecole d’été de Probabilités de Saint-Flour 2003. Lecture Notes in Mathematics, Springer.
- [54] Miermont, Grégory. Self-similar fragmentations derived from the stable tree. I. Splitting at heights. *Probab. Theory Related Fields* **127** (2003), no. 3, 423–454.
- [55] Pitman, Jim. Cyclically stationary Brownian local time processes. *Probab. Theory Related Fields* **106** (1996), no. 3, 299–329.
- [56] Pitman, J. *Combinatorial stochastic processes*. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Lecture Notes in Mathematics, 1875. Springer-Verlag, Berlin, 2006.
- [57] Pitman, Jim ; Yor, Marc. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** (1997), no. 2, 855–900.
- [58] Propp, James Gary ; Wilson, David Bruce. How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. 7th Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996). *J. Algorithms* **27** (1998), no. 2, 170–217.
- [59] Rémy, Jean-Luc. Un procédé itératif de dénombrement d’arbres binaires et son application à leur génération aléatoire. *RAIRO Inform. Théor.* **19** (1985), no. 2, 179–195.
- [60] Reynaud-Bouret, P. (2003). Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Related Fields* **126**, 103–153.
- [61] Ruelle, D. (1987) A mathematical reformulation of Derrida’s REM and GREM. *Commun. Math. Phys.* **108**, 225–239.
- [62] Rüschenendorf, L. and Woerner, J. Expansion of transition distributions of Lévy processes in small time. *Bernoulli*, **8** :81–96, 2002.

- [63] Spitzer, Frank. *Principles of random walks*. Second edition. Graduate Texts in Mathematics, Vol. 34. Springer-Verlag, New York-Heidelberg, 1976.
- [64] Sudakov V.N. and Tsirel'son B.S. Extremal properties of half-spaces for spherically invariant measures. *Zap. Nauch. Sem. LOMI* **41** (1974), 14–24. English translation in : *J. Soviet Math.* **9** (1978), 9–18.
- [65] Talagrand, M. (1991). A new isoperimetric inequality for product measure, and the concentration of measure phenomenon. Israel Seminar (GAFA). *Lecture Notes in Math.* **1469**, 91–124. Springer, Berlin.
- [66] Viennot, Gérard Xavier. Heaps of pieces. I. Basic definitions and combinatorial lemmas. *Combinatoire énumérative*, 321–350, Lecture Notes in Math., 1234, Springer, Berlin, 1986.
- [67] Vigon, Vincent. Votre Lévy rampe-t-il? *J. London Math. Soc. (2)* **65** (2002), no. 1, 243–256.
- [68] Zolotarev, V. M. *One-dimensional stable distributions*. Translations of Mathematical Monographs, **65**. American Mathematical Society, Providence, RI, 1986.

Chapitre 3

Almost sure convergence of the simple random walk to Brownian motion

We give a recursive algorithm generating a sequence of simple random walks on \mathbb{Z} , converging almost surely to a Brownian motion. One obtains by the same method conditional versions of the simple random walk converging almost surely to the excursion, the bridge, the meander or the normalized pseudobridge. Connections with random trees, urn processes and Pitman's Chinese restaurant are discussed.

3.1 Introduction

It is one of the most basic facts in probability theory that random walks, after proper rescaling, converge to Brownian motion. However, Donsker's classical theorem [7] only states a convergence in law. Various results of almost sure convergence exist (see [19, 11, 6, 13, 14] among others, or the survey [5]) but involve rather intricate relations between the converging sequence of random walks and the limiting Brownian motion.

Specific methods exist for the simple random walk on \mathbb{Z} (see for instance [12]) but they fail to construct random walks with prescribed length n for each n . The aim of this paper is to give a new result of almost sure convergence with an explicit and simple "recipe" to construct a converging sequence of simple random walks. The algorithm is described in the next section and leads to the following

Theorem 1 *There exist on a common probability space a family $(S^n, n \geq 1)$ of random walks on \mathbb{Z} and a linear Brownian motion $(B_t, 0 \leq t \leq 1)$ such that :*

(i) for every n , S^n has the law of a simple random walk with n steps starting at 0,

(ii) almost surely, one has the uniform convergence

$$\left(\frac{S_{[nt]}^n}{\sqrt{n}}, t \in [0, 1] \right) \xrightarrow{U} (B_t, t \in [0, 1])$$

Moreover, there exists a family $(U_n, n \geq 1)$ of independent random variables such that U_1 is a standard Bernoulli random variable, U_n is uniformly distributed on $\{1, 2, \dots, 4n - 4\}$ for every $n \geq 2$, and such that S^{2n-1} is generated by (U_1, \dots, U_n) .

An important feature of our construction is that it can be adapted to conditional versions of the random walk, yielding the following generalization :

Theorem 2 *There exist on a common probability space a family $(S^n, n \geq 1)$ of random walks on \mathbb{Z} and a process $(\tilde{B}_t, t \in [0, 1])$ such that for every n , S^n has the law, respectively, of*

- (1) a random walk of length $2n$ starting at 0 and conditioned to return to 0 at time $2n$,
- (2) a random walk of length $2n$ starting at 0, conditioned to return to 0 at time $2n$ and to stay positive from time 1 to $2n - 1$,
- (3) a random walk of length n starting at 0 and conditioned to stay positive from time 1 to $n - 1$,

and such that almost surely, one has the uniform convergence

$$\left(\frac{S_{c_n(t)}^n}{\sqrt{n}}, t \in [0, 1] \right) \xrightarrow{U} (\tilde{B}_t, t \in [0, 1])$$

where where $c_n(t)$ equals respectively $[2nt]$, $[2nt]$, $[nt]$ and \tilde{B} is respectively :

- (1) a Brownian bridge,
- (2) a Brownian excursion,
- (3) a Brownian meander.

An analogous result holds for random walks biased by their local time at 0 and converging almost surely to the Brownian pseudobridge [4].

Theorem 3 *There exist on a common probability space a family $(S^n, n \geq 1)$ of random walks on \mathbb{Z} and a process $(\hat{B}_t, t \in [0, 1])$ such that*

- (a) for every n , S^n has length $2n$, starts at 0 and ends at 0,
- (b) for every path \mathcal{P} of length $2n$ starting at 0 and ending at 0, the probability $\mathbb{P}(S^n = \mathcal{P})$ is proportional to $1/L^0(\mathcal{P})$ where $L^0(\mathcal{P})$ is the number of visits of 0 of \mathcal{P} ,
- (c) almost surely, one has the uniform convergence

$$\left(\frac{S_{[2nt]}^n}{\sqrt{n}}, t \in [0, 1] \right) \xrightarrow{U} (\hat{B}_t, t \in [0, 1])$$

where \hat{B} is a normalized Brownian pseudobridge.

From an algorithmic point of view, our algorithms construct S^n recursively and are therefore interruptible. Notice that the natural method to construct a discrete version of the Brownian pseudobridge would consist in running a random walk up the N -th return time to 0, for large N , and then rescaling. But this method has the major drawback that $E(T_N) = \infty$.

It is well-known in combinatorics that there exists natural bijections between binary trees and excursions of the simple random walk. In this context, the construction method we shall describe in case (2) of Theorem 2 is a pathwise counterpart of the tree-generating algorithm described in [17].

Finally, let us mention that the convergence rate for our construction can be bounded by

$$\mathbb{E} \left[\int_0^1 \left| \frac{S^n_{[tn]}}{\sqrt{n}} - B_t \right| dt \right] \leq \frac{c}{n^{1/4}} \quad (3.1)$$

for some constant $c > 0$. This indicates a slower convergence rate than the optimal one, which is $O(\log n/\sqrt{n})$ and is achieved in [13].

The algorithms generating the converging families of random walks (S^n) are described in the next section. Theorems 1 and 2 are proved in Sections 3 and 4. Further results and the proof of Theorem 3 are given in Section 5.

3.2 Description of the algorithms

Recall that for paths of the simple random walk, an excursion is a part of a path between two consecutive zeros, the bridge is the part before the last zero and the meander is the part after the last zero. The basic algorithm generates random excursions. The other algorithms are generalizations of the first one.

3.2.1 The case of the excursion

We call a hat a sequence of a positive step followed by a negative one. First let us describe the two basic operations that will enable us to extend our paths.

Inserting a sequence of steps A in the path \mathcal{P} at time t means concatenating

- the part of \mathcal{P} before time t
- the sequence A
- the part of \mathcal{P} after time t .

If $T \in \mathbb{N}$ and \mathcal{P} is a path starting at 0 and such that $\mathcal{P}(T) > 0$, *lifting the Dyck path before time T* means the following. Let

$$T' = 1 + \sup\{n \leq T, \mathcal{P}(n) < \mathcal{P}(T)\}$$

Then form the new path \mathcal{P}' by inserting a positive step at time T' and then a negative step at time $T + 1$. Remark that if $\mathcal{P}(T - 1) = \mathcal{P}(T) - 1$, then $T' = T$ and lifting the Dyck path before time T amounts to inserting a hat at time T .

The algorithm to generate the $2n$ -step excursion \mathcal{S}^n is the following. Start with \mathcal{S}^1 a hat. Then recursively, to generate \mathcal{S}^{n+1} from \mathcal{S}^n :

FIG. 3.1 – Inserting a hat at time T

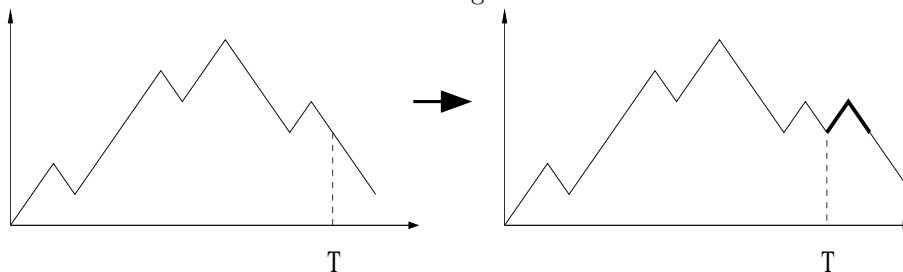
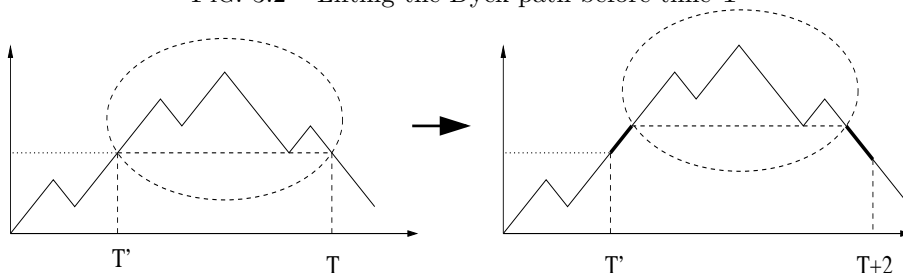


FIG. 3.2 – Lifting the Dyck path before time T



- I. Choose a random time t uniformly on $\{1, \dots, 2n - 1\}$.
- II. with probability $1/2$ insert a positive hat at time t .
- III. with probability $1/2$ lift the Dyck path before time t .

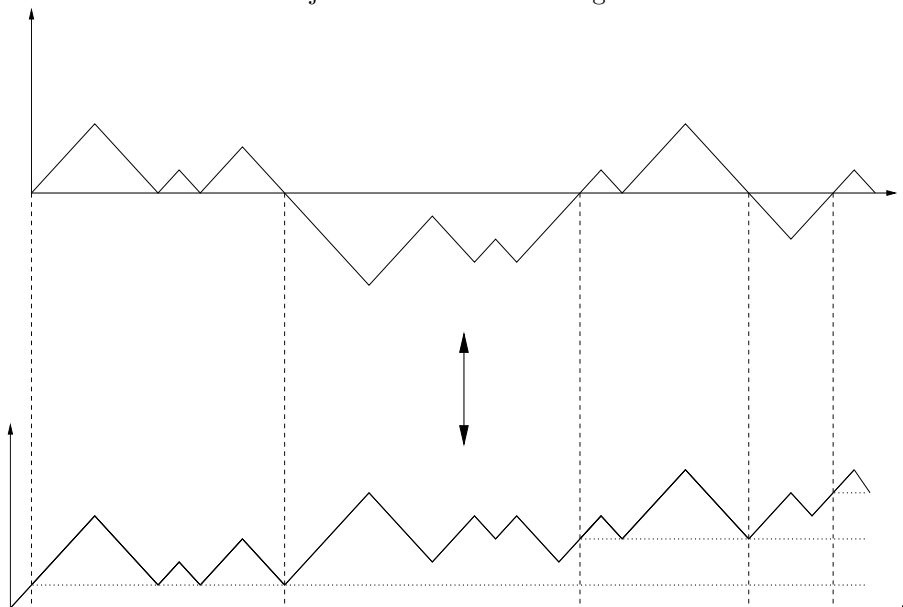
Of course, all the choices are assumed independent.

3.2.2 The bridge and the meander

To construct a bridge \mathcal{S}^n of length $2n$, start with \mathcal{S}^0 the empty path. Then recursively, to generate \mathcal{S}^{n+1} from \mathcal{S}^n :

- I. Choose a random time t uniformly on $\{0, \dots, 2n\}$.
- II. If $\mathcal{S}^n(t) = 0$, insert a positive or negative hat at time t with respective probabilities $1/2-1/2$.
- III. If $\mathcal{S}^n(t) > 0$,
 1. with probability $1/2$ insert a positive hat at time t .

FIG. 3.3 – The bijection between the bridge and the meander



2. with probability $1/2$ lift the Dyck path before time t .

Of course if $\mathcal{S}^n(t) < 0$, the operation is similar. This algorithm is nothing but a generalization of the previous one, the difference being that one may choose a time t at which $\mathcal{S}^n(t) = 0$, enabling to add new excursions.

The meander-generating algorithm is based on the classical bijection between the bridge and the meander. To construct a meander of length $2n + 1$ from a bridge of length $2n$, replace each negative excursion of the bridge by its symmetric positive excursion and replace the last negative step of this symmetric positive excursion by a positive step. Finally, add a first positive step to the whole path. See the figure.

This is the discrete analogue of the fact (see for instance in [18]) that if $(B_t, 0 \leq t \leq 1)$ is a Brownian bridge, then $(|B_t| + L_t, 0 \leq t \leq 1)$ is a Brownian meander, L_t standing for the local time at 0 up to time t .

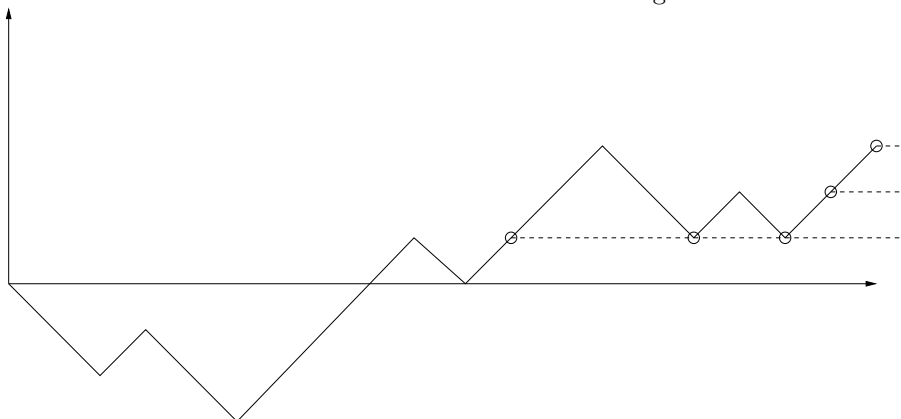
Using the bijection, the algorithm for the bridge translates into an algorithm for the meander.

3.2.3 The general case

Since we know how to construct the bridge and the meander, constructing the whole path is just a matter of combining the two algorithms. For the sake of completeness, we want to give an explicit description here.

This requires to define the notion of “visiblens from the right”. A point t

FIG. 3.4 – Points visible from the right



in the positive meander of a path \mathcal{P} is visible from the right if

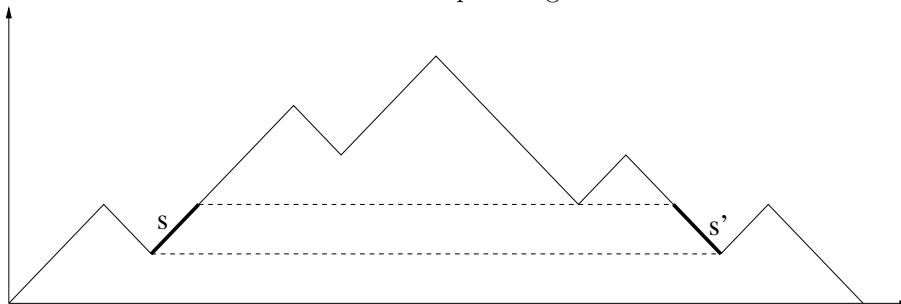
$$\mathcal{P}(t) = \min\{\mathcal{P}(n), n \geq t\}.$$

In the language of fluctuation theory, t is a weak ascending ladder time of the dual path. See the figure, where the circled points are those which are visible from the right.

Our algorithm generates $\mathcal{S}^n := S^{2n-1}$. One constructs S^{2n} by adding a last random step to \mathcal{S}^n . First choose \mathcal{S}^1 at random. Then to construct \mathcal{S}^{n+1} from \mathcal{S}^n :

- I. Choose a random time t uniformly on $\{0, \dots, 2n-1\}$.
- II. If $\mathcal{S}^n(t) = 0$, insert a positive or negative hat at time t with respective probabilities $1/2-1/2$.
- III. If t is in a positive excursion,
 1. with probability $1/2$ insert a positive hat at time t .
 2. with probability $1/2$ lift the Dyck path before time t .
- IV. If t is in the meander of \mathcal{S}^n and if this meander is positive,
 1. If t is invisible from the right, or if t is visible from the right and $\mathcal{S}^n(t)$ is even, proceed as in III.
 2. If t is visible from the right and $\mathcal{S}^n(t)$ odd,
 - a. with probability $1/2$ insert at time t a positive hat,

FIG. 3.5 – Two steps facing each other



b. with probability $1/2$ insert at time t two positive steps.

If t is in a negative excursion or in a negative meander, the procedure is the exact analogue of **III** or **IV**.

At this point, for each algorithm we have to prove two points. First, for each n , the algorithm at time n yields a random excursion, bridge or path with the uniform distribution. Second, there is uniform convergence almost surely.

The proof of the first point is given for the last algorithm. The proof for the other cases is completely similar, indeed simpler. In turn, the proof of the second point is given for the excursion. The proof for the other cases would be analogous, but the notations would be heavier.

3.3 The combinatorial part of Theorem 1

We first introduce some more definition. A positive step $s = (\mathcal{P}_n, \mathcal{P}_{n+1})$ of a path \mathcal{P} is visible from the right if

$$\mathcal{P}_{n+1} = \inf\{\mathcal{P}_k, k \geq n + 1\}$$

Remark that this entails that s is in the meander and that the meander is positive.

With the same notations, if s is not visible from the right and s is in the meander or in a positive excursion, the negative step facing s is the step $s' = (\mathcal{P}_m, \mathcal{P}_{m+1})$ such that

$$m + 1 = \inf\{i > n, \mathcal{P}_i = \mathcal{P}_n\}$$

See the figure. If the meander is negative, the definition of a time or of a step “visible from the right” is similar.

We want to prove by induction on n that for every n , \mathcal{S}^n is uniformly distributed over all possible paths of length $2n - 1$. This is trivial for $n = 1$. Now assume that the induction hypothesis holds for n . Given a path \mathcal{P} of length $2n + 1$, we would like to compute $\mathbb{P}(\mathcal{S}^{n+1} = \mathcal{P})$. To do so, let us first define

a coloured path as a path where one step s is coloured red with the following conditions :

(P1) If s is positive (resp. negative), then s is in a positive (resp. negative) excursion or in a positive (resp. negative) meander.

(P2) If s is visible from the right, then the left extremity of s is at an odd altitude.

Observe that given a path of length $2n+1$, there are exactly n ways to colour a step so as to obtain a coloured path satisfying (P1) and (P2). Also observe that if after each step of the algorithm we colour exactly one step s , namely

- in the cases **II**, **III 1** or **IV 2**, s is the first inserted step,
- in the case **III 2**, s is the step preceding the first inserted step,

then one obtains a coloured path satisfying (P1) and (P2). Call this path $Col(\mathcal{S}_n)$ for each n . Given a path \mathcal{P} of length $2n+1$, denote by $C(\mathcal{P})$ the set of coloured versions of \mathcal{P} . Then

$$|C(\mathcal{P})| = n \tag{3.2}$$

Remark. Suppose that the time t chosen by the algorithm to construct \mathcal{S}^{n+1} from \mathcal{S}^n is such that $0 \leq \mathcal{S}_{t-1}^n < \mathcal{S}_t^n$. Then as mentioned before, from the point of view of the non-coloured path, **III 1** and **III 2** are equivalent : both insert a hat at time t . However, **III 1** and **III 2** yield two different coloured paths, the coloured step being different in the two cases.

Conversely, the knowledge of $Col(\mathcal{S}_n)$ enables us to recover \mathcal{S}_{n-1} . Indeed, suppose that the coloured step s is positive. Let s' be the step following s . Then

- If s' is negative, \mathcal{S}_{n-1} is obtained from $Col(\mathcal{S}_n)$ by removing s and s' (cases **II**, **III 1**, **IV 2 a**).
- If s' is positive and not visible from the right, \mathcal{S}_{n-1} is obtained from $Col(\mathcal{S}_n)$ by removing s' and the negative step facing s' (case **III 2**).
- If s' is positive and visible from the right, \mathcal{S}_{n-1} is obtained from $Col(\mathcal{S}_n)$ by removing s and s' (case **IV 2 b**).

When s is a negative step, the method to recover \mathcal{S}_{n-1} is similar.

As a consequence, given a coloured path \mathcal{Q} of length $2n+1$, there exists a unique path $h(\mathcal{Q})$ of length $2n-1$ producing \mathcal{Q} by the algorithm of Theorem 1. Moreover,

$$\mathbb{P}(Col(\mathcal{S}_{n+1}) = \mathcal{Q} | \mathcal{S}_n = h(\mathcal{Q})) = \frac{1}{4n} \tag{3.3}$$

since conditionally on $\mathcal{S}_n = h(\mathcal{Q})$, we have $4n$ equiprobable possibilities to extend $h(\mathcal{Q})$, each possibility generating a different coloured path.

Now given a path \mathcal{P} of length $2n+1$, we can compute :

$$\begin{aligned}
\mathbb{P}(\mathcal{S}_{n+1} = \mathcal{P}) &= \sum_{\mathcal{Q} \in C(\mathcal{P})} \mathbb{P}(\text{Col}(\mathcal{S}_{n+1}) = \mathcal{Q}) \\
&= \sum_{\mathcal{Q} \in C(\mathcal{P})} \mathbb{P}(\text{Col}(\mathcal{S}_{n+1}) = \mathcal{Q} | \mathcal{S}_n = h(\text{Col}(\mathcal{S}_{n+1}))) \mathbb{P}(\mathcal{S}_n = h(\text{Col}(\mathcal{S}_{n+1}))) \\
&= \sum_{\mathcal{Q} \in C(\mathcal{P})} \mathbb{P}(\text{Col}(\mathcal{S}_{n+1}) = \mathcal{Q} | \mathcal{S}_n = h(\text{Col}(\mathcal{S}_{n+1}))) \cdot 2^{1-2n} \\
&= \sum_{\mathcal{Q} \in C(\mathcal{P})} \frac{2^{1-2n}}{4n} \\
&= |C(\mathcal{P})| \cdot \frac{2^{1-2n}}{4n} \\
&= 2^{-2n-1}
\end{aligned}$$

where the third equality follows from the induction hypothesis, the fourth equality from (3.3) and the last equality from (3.2).

3.4 Proof of almost sure convergence

Recall that we are dealing with the excursion. The path with length $2n$ is denoted by \mathcal{S}^n and its normalized version by $\widehat{\mathcal{S}}^n$: for $t \in [0, 1]$,

$$\widehat{\mathcal{S}}_t^n = \frac{\mathcal{S}_{[2nt]}^n}{\sqrt{2n}}$$

where $[2nt]$ stands for the integer part of $2nt$. Recall that by Donsker's theorem, the discrete excursion converges in distribution to the continuous excursion.

3.4.1 Moving steps and local time

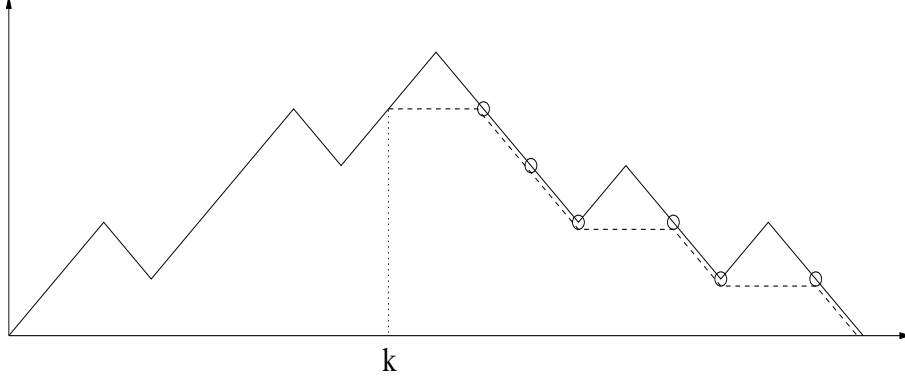
Let $(\mathcal{S}_k^n, \mathcal{S}_{k+1}^n)$ be a step of \mathcal{S}^n inserted at time n . We associate with this step a family s of steps with exactly one step in each path \mathcal{S}^m , $m \geq n$. This family is defined by induction : if the step of s in \mathcal{S}^m is $(\mathcal{S}_{k_m}^m, \mathcal{S}_{k_m+1}^m)$ then the step of s in \mathcal{S}^{m+1} is :

- $(\mathcal{S}_{k_m}^{m+1}, \mathcal{S}_{k_m+1}^{m+1})$ if \mathcal{S}^{m+1} is obtained from \mathcal{S}^m by inserting two steps on the right of k_m
- $(\mathcal{S}_{k_m+2}^{m+1}, \mathcal{S}_{k_m+3}^{m+1})$ if \mathcal{S}^{m+1} is obtained from \mathcal{S}^m by inserting two steps on the left of k_m
- $(\mathcal{S}_{k_m+1}^{m+1}, \mathcal{S}_{k_m+2}^{m+1})$ if \mathcal{S}^{m+1} is obtained from \mathcal{S}^m by inserting one step on the left of k_m and one step on the right of k_m .

We call such a family s a *moving step*. We shall denote $x_m(s) = k_m$ and $y_m(s) = \mathcal{S}_{k_m}^m$. Define the set $E_n(k)$ as the set of times $j \in [k+1, 2n-1]$ such that

$$\mathcal{P}(j) = \min\{\mathcal{P}(i), i \in [k, j]\}$$

FIG. 3.6 – The set $E_n(k)$



See Figure 6, where the circled points are those which belong to $E_n(k)$. The local time for k in \mathcal{S}^n is defined by $L_n(k) = |E_n(k)|$.

3.4.2 Martingale properties

Let s be a moving step. We shall use the abridged notation $L_n = L_n(x_n(s))$. A key remark is that if the random time t chosen by the algorithm to construct \mathcal{S}^{n+1} from \mathcal{S}^n is in $E_n(x_n(s))$, then $L_{n+1} - L_n = 1$. Otherwise, if $t \notin E_n(x_n(s))$, $L_{n+1} = L_n$.

As a consequence, denoting

$$\frac{a_{n+1}}{a_n} = 1 + \frac{1}{2n-1}$$

so that $a_n \sim c\sqrt{n}$, and

$$M_n(s) = \frac{L_n}{a_n}$$

we check that $(M_n(s))$ is a positive martingale, and therefore converges almost surely :

$$\begin{aligned} & \mathbb{E} (M_{n+1}(s) | M_n(s)) \\ &= \frac{L_n}{2n-1} \left(\frac{L_n+1}{a_{n+1}} \right) + \frac{2n-1-L_n}{2n-1} \left(\frac{L_n}{a_{n+1}} \right) \\ &= \frac{a_n M_n(s)}{2n-1} \left(\frac{a_n M_n(s) + 1}{a_{n+1}} \right) + \frac{2n-1-a_n M_n(s)}{2n-1} \left(\frac{a_n M_n(s)}{a_{n+1}} \right) \\ &= \frac{a_n M_n(s)}{a_{n+1}} \left(1 + \frac{1}{2n-1} \right) = M_n(s) \end{aligned}$$

On the other hand, setting $x'_n(s) = x_n(s)/2n$ we have

$$\begin{aligned} & \mathbb{E} (x'_{n+1}(s)|x'_n(s), M_n(s)) \\ &= \left(\frac{x_n}{2n-1}\right) \left(\frac{x_n+2}{2n+2}\right) + \left(\frac{2n-1-x_n}{2n-1}\right) \left(\frac{x_n}{2n+2}\right) + O\left(\frac{L_n}{2n} \cdot \frac{1}{n}\right) \\ &= x'_n(s) + O\left(\frac{M_n(s)}{n^{3/2}}\right) \end{aligned}$$

which entails that $x'_n(s)$ converges almost surely, as well.

3.4.3 L^2 estimates

We have the following estimates :

$$\begin{aligned} & \mathbb{E} ((M_{n+1}(s) - M_n(s))^2 | M_n(s)) \\ &= \frac{L_n}{2n-1} \left(\frac{L_n+1}{a_{n+1}} - \frac{L_n}{a_n}\right)^2 + \frac{2n-1-L_n}{2n-1} \left(\frac{L_n}{a_{n+1}} - \frac{L_n}{a_n}\right)^2 \\ &\leq \frac{L_n}{2n-1} \left(\frac{1}{a_{n+1}}\right)^2 + \left(\frac{L_n}{a_n}\right)^2 \left(\frac{a_n}{a_{n+1}} - 1\right)^2 \\ &\leq c_1 \frac{M_n(s)}{n\sqrt{n}} + c_2 \frac{M_n(s)^2}{n^2} \end{aligned}$$

where c_1, c_2 are universal constants which do not depend on s . Using a maximal inequality we obtain

$$\mathbb{E}(\sup_{n \geq N} (M_N(s) - M_n(s))^2) \leq \frac{c_3 M_N(s)}{\sqrt{N}} \quad (3.4)$$

where again c_3 does not depend on s . Similarly,

$$\begin{aligned} & \mathbb{E} ((x'_{n+1}(s) - x'_n(s))^2 | x'_n(s), M_n(s)) \\ &\leq \left(\frac{x_n}{2n-1}\right) \left(\frac{x_n+2}{2n+2} - \frac{x_n}{2n}\right)^2 + \left(\frac{2n-1-x_n}{2n-1}\right) \left(\frac{x_n}{2n+2} - \frac{x_n}{2n}\right)^2 + O\left(\frac{L_n}{2n} \cdot \frac{1}{n^2}\right) \\ &\leq \frac{c_4}{n^2} \end{aligned}$$

and

$$\mathbb{E}(\sup_{n \geq N} (x'_N(s) - x'_n(s))^2) \leq \frac{c_5}{N} \quad (3.5)$$

3.4.4 The discrete skeleton

The key argument for the proof of strong convergence is the following : for every moving step s ,

$$\left(\frac{x_n(s)}{2n}, \frac{\mathcal{S}_{x_n(s)}^n}{\sqrt{2n}}\right)$$

converges almost surely as $n \rightarrow \infty$ to some point $(X(s), Y(s))$.

Recall that the convergence of $x'_n(s) = x_n(s)/2n$ was established in Section 4.2. Remark that $x_n(s)$ looks like a Polya urn. The time t chosen by the algorithm to construct \mathcal{S}^{n+1} from \mathcal{S}^n corresponds to the ball drawn from the urn, the two colours corresponding to t being on the left or on the right of the step of s in \mathcal{S}^n . In general, if t is on the left of s , one adds two balls to the left of s and if t is on the right of s , one adds two balls to the right of s . The difference with a Polya urn is that if $t \in E_n(x_n(s))$, one adds one ball to the left and one to the right.

On the other hand, as the martingale $M_n(s)$ introduced in Section 4.2 converges almost surely and as $a_n \sim c\sqrt{n}$, the quantity $L_n/\sqrt{2n}$ converges almost surely to some real $Y'(s)$. Moreover, each time L_n increases, $\mathcal{S}_{x_n(s)}^n$ either increases (the algorithm lifts a Dyck path) with probability $1/2$ or remains constant (the algorithm inserts a hat) with probability $1/2$, independently of the past. That is, $\mathcal{S}_{x_n(s)}^n$ is the sum of L_n independent, standard Bernoulli variables and by the law of large numbers, $\mathcal{S}_{x_n(s)}^n/\sqrt{2n} \rightarrow Y'(s)/2$ almost surely.

Thus we get a discrete ‘‘skeleton’’ $\cup_s \{(X(s), Y(s))\}$, where the union is over all moving steps. If we denote by \mathbf{B} the closure in \mathbb{R}^2 of $\cup_s \{(X(s), Y(s))\}$, then the curves $\widehat{\mathcal{S}}^n$ converge in some sense to \mathbf{B} and it is intuitively clear that \mathbf{B} is the graph of a Brownian excursion. However, a rigorous proof requires some technical arguments, which are presented in the next two subsections.

Let us begin with some estimates. We want to bound

$$I_n = \mathbb{E} \left[\int_0^1 d((t, \widehat{\mathcal{S}}_t^n), \mathbf{B}) dt \right]$$

where $d(., .)$ stands for the euclidean distance in \mathbb{R}^2 . We have

$$I_n = \sum_{k=1}^{2n} \mathbb{E} \left[\int_{(k-1)/2n}^{k/2n} d((t, \widehat{\mathcal{S}}_t^n), \mathbf{B}) dt \right]$$

and for every $1 \leq k \leq 2n$,

$$\int_{(k-1)/2n}^{k/2n} d((t, \widehat{\mathcal{S}}_t^n), \mathbf{B}) dt \leq \int_{(k-1)/2n}^{k/2n} d((t, \widehat{\mathcal{S}}_t^n), (X(s_t), Y(s_t))) dt$$

where s_t is the moving step containing the step $(\mathcal{S}_{k-1}^n, \mathcal{S}_k^n)$ in \mathcal{S}^n . The estimates (3.4) and (3.5) entail that for every $t \in [(k-1)/n, k/n]$, conditionally on $L^n(s_t)$,

$$\mathbb{E} \left[d((t, L^n(s_t)/\sqrt{2n}), (X(s), Y'(s))) | L^n(s_t) \right] \leq \frac{c_6 L^n(s_t)}{n^{3/4}}$$

Using the fact that \mathcal{S}_t^n is the sum of $L^n(s_t)$ independent Bernoulli random variables, one can replace $L^n(s_t)/\sqrt{2n}$ by $\widehat{\mathcal{S}}_t^n$ and $Y'(s)$ by $Y'(s)/2 = Y(s)$ in the last inequality without changing the order of magnitude of the upper bound :

$$\mathbb{E} \left[d((t, \widehat{\mathcal{S}}_t^n), (X(s), Y(s))) | L^n(s_t) \right] \leq \frac{c_7 L^n(s_t)}{n^{3/4}}$$

Integrating between $(k-1)/2n$ and $k/2n$ yields

$$\mathbb{E} \left[\int_{(k-1)/2n}^{k/2n} d((t, \widehat{\mathcal{S}}_t^n), \mathbf{B}) dt | L^n(k) \right] \leq \frac{c_7}{2n} \cdot \frac{L^n(k)}{n^{3/4}}$$

Summing over k and integrating with respect to the law of $\widehat{\mathcal{S}}^n$,

$$I_n \leq c_7 \sup_k \mathbb{E} \left[\frac{L^n(k)}{n^{3/4}} \right] = 2c_7 \sup_k \mathbb{E} \left[\frac{S_k^n}{n^{3/4}} \right] \leq c_8 \mathbb{E} \left[\sup_t \frac{\widehat{\mathcal{S}}_t^n}{n^{1/4}} \right] \leq \frac{c_9}{n^{1/4}} \quad (3.6)$$

since by Donsker's theorem, $\mathbb{E}[\sup_t(\widehat{\mathcal{S}}_t^n)]$ is bounded in n . Remark that (3.6) entails that the set $\widetilde{X} = \{X(s)\}$ is almost surely dense in $[0, 1]$.

3.4.5 \mathbf{B} is the graph of a continuous function

We want to prove that with probability 1, \mathbf{B} is the graph of a continuous function $(B_t, 0 \leq t \leq 1)$. First consider a property Q defined on renormalized trajectories. We have for every integer N ,

$$\mathbb{P}(\forall n \geq N, Q(\widehat{\mathcal{S}}^n)) \leq \mathbb{P}(Q(\widehat{\mathcal{S}}^N))$$

and consequently

$$\mathbb{P}(\exists N_0, \forall n \geq N_0, Q(\widehat{\mathcal{S}}^n)) \leq \liminf_n \mathbb{P}(Q(\widehat{\mathcal{S}}^n)) \quad (3.7)$$

We apply this to the crossings of an interval. If $q < q'$ are two rational numbers and f is a function $[0, 1] \rightarrow \mathbb{R}$, let

$$Cros(f, q, q') = \inf_{f(t) < q, f(t') > q'} |t - t'|$$

By Donsker's theorem, $Cros(\widehat{\mathcal{S}}^n, q, q')$ converges in distribution, as $n \rightarrow \infty$, to $Cros(\beta, q, q')$ where β is a Brownian excursion. In particular, for every integer M set

$$E_{q, q'}(M) = \{\exists N_0, \forall n \geq N_0, Cros(\widehat{\mathcal{S}}^n, q, q') \leq \frac{1}{M}\}$$

Then using (3.7), one obtains that the event

$$E_{q, q'} = \bigcap_{M \in \mathbb{N}} E_{q, q'}(M)$$

has probability 0.

Let E be the event that \mathbf{B} is discontinuous, more precisely, that there exist a sequence (s_k) of moving steps and a real $t \in [0, 1]$ such that $X(s_k) \rightarrow t$ and $\liminf(Y(s_k)) < \limsup(Y(s_k))$. Let q, q' be two rational numbers satisfying

$$\liminf(Y(s_k)) < q < q' < \limsup(Y(s_k))$$

Then for every $M \in \mathbb{N}$, there exist two moving steps $s'_1, s'_2 \in \{s_k, k \in \mathbb{N}\}$ such that $|X(s'_1) - X(s'_2)| \leq 1/2M$ and $Y(s'_1) < q, Y(s'_2) > q'$.

For n sufficiently large, if the step of s'_1 in \mathcal{S}^n is $(\mathcal{S}_k^n, \mathcal{S}_{k+1}^n)$, let $t_1(n) = k/2n$ and define likewise $t_2(n)$. Then since the moving steps converge to their limits in \mathbb{R}^2 , for n sufficiently large, $\widehat{\mathcal{S}}_{t_1(n)}^n < q, \widehat{\mathcal{S}}_{t_2(n)}^n > q'$ and $|t_1(n) - t_2(n)| \leq 1/M$. Hence for n sufficiently large, $\text{Cros}(\widehat{\mathcal{S}}^n, q, q') \leq 1/M$ or equivalently, $E_{q,q'}(1/M)$ occurs. As this is true for every positive integer M , $E_{q,q'}$ occurs. We deduce that $E \subset \cup_{q,q'} E_{q,q'}$ and thus $\mathbb{P}(E) = 0$.

Recall that almost surely, \widetilde{X} is dense in $[0, 1]$. Moreover, it is clear that on the event $\{\widetilde{X} \text{ is dense}\} \cap E^c$, the set \mathbf{B} is the graph of a continuous function $(B_t, 0 \leq t \leq 1)$. As a consequence of Donsker's theorem, $(B_t, 0 \leq t \leq 1)$ is a Brownian excursion. Since a Brownian excursion is almost surely $(1/2 - \epsilon)$ -Hölder-continuous for every $\epsilon > 0$, (3.1) follows easily from (3.6).

3.4.6 Uniform convergence to \mathbf{B}

So far, we have established a convergence of $\widehat{\mathcal{S}}^n$ to B in the sense of convergence of points in \mathbb{R}^2 . We would like to obtain a more orthodox convergence, namely the almost sure, uniform convergence of $(\widehat{\mathcal{S}}_t^n, t \in [0, 1])$ to $(B_t, t \in [0, 1])$. Of course, if we could show directly that $(\widehat{\mathcal{S}}_t^n, t \in [0, 1])$ converges uniformly to some function $(B_t, t \in [0, 1])$, the previous subsection would be useless. However, our proof uses the existence and the almost sure continuity of $(B_t, t \in [0, 1])$.

First consider two rectangles $I \times J$ and $I' \times J'$, where I, I', J, J' are intervals of $[0, 1]$, with $\inf(I) > \inf(I'), \sup(I) < \sup(I'), \inf(J) > \inf(J'), \sup(J) < \sup(J')$.

Define n_0 as the first integer such that the graph of $\widehat{\mathcal{S}}^{n_0}$ intersects $I \times J$, u_0 as the first step of $\widehat{\mathcal{S}}^{n_0}$ intersecting $I \times J$ and s_0 as the corresponding moving step. Let n'_0 be the first integer $n \geq n_0$ such that the step of s_0 in $\widehat{\mathcal{S}}^n$ does not intersect $I' \times J'$.

Then by induction let n_1 be the first integer $\geq n'_0$ such that $\widehat{\mathcal{S}}^{n_1}$ intersects $I \times J$, u_1 the first step of $\widehat{\mathcal{S}}^{n_1}$ intersecting $I \times J$ and s_1 the corresponding moving step, n'_1 the first integer $n \geq n_1$ such that the step of s_1 in $\widehat{\mathcal{S}}^n$ does not intersect $I' \times J'$ and so on.

Of course for a given integer k , n_k or n'_k may not exist, in which case the sequence (n_0, n_1, \dots) is finite. Let $E(I, I', J, J')$ be the event that the sequence (n_0, n_1, \dots) is infinite. From the estimate (3.4) and Chebyshev's inequality, it appears that $E(I, I', J, J')$ has probability 0.

Let E be the event that $(\widehat{\mathcal{S}}_t^n, t \in [0, 1])$ does not converge uniformly to $(B_t, t \in [0, 1])$. On E there exists an $\varepsilon > 0$ such that infinitely often, $\|\widehat{\mathcal{S}}^n - B\|_\infty > \varepsilon$. Since $(B_t, t \in [0, 1])$ is almost surely bounded, conditionally on E there exists almost surely an x such that for infinitely many n , one can find t_n satisfying $\widehat{\mathcal{S}}_{t_n}^n > x + \varepsilon, B_{t_n} < x$ (or conversely, $\widehat{\mathcal{S}}_{t_n}^n < x, B_{t_n} > x + \varepsilon$).

Moreover, since the set of such t_n is bounded, and since B is continuous, there exist almost surely two nontrivial intervals $I \subset I' \subset [0, 1]$, with $\inf(I) > \inf(I')$, $\sup(I) < \sup(I')$, such that $B_t < x$ for $t \in I'$ and that for infinitely many n , one can find $t_n \in I$ satisfying $\widehat{\mathcal{S}}_{t_n}^n > x + \varepsilon$.

Consider the intervals $J = [x + \varepsilon, x + M]$ and $J' = [x, \infty)$. Then for $M > 0$ sufficiently large, one can find infinitely many $t_n \in I$ satisfying $\widehat{\mathcal{S}}_{t_n}^n \in [x + \varepsilon, x + M]$. In other words, for such a real M , infinitely often, $\widehat{\mathcal{S}}^n$ intersects the rectangle $I \times J$. Moreover, since \mathbf{B} does not intersect the rectangle $I' \times J'$, every moving step intersecting $I \times J$ eventually exits from $I' \times J'$. Therefore the event $E(I, I', J, J')$ occurs.

So we have proved that on E , almost surely, the union $\cup_{I, I', J, J'} E(I, I', J, J')$ occurs. As we may restrict the union to intervals with rational coordinates, we deduce that on E , almost surely, a countable union of negligible sets occurs. Hence E itself is negligible.

3.5 Further results

3.5.1 Trees

We deal here with the excursion-generating algorithm. As already mentioned, our algorithm is the exact counterpart of an algorithm due to Rémy [17] generating random binary trees, the link between binary trees and excursions of the random walk being the famous Lukasiewicz correspondence between trees and Lukasiewicz paths (see for instance [9] and the references therein).

Alternatively, it is well-known that Brownian excursions are naturally associated to continuous trees. In particular, sampling n points uniformly at random in a Brownian excursion is equivalent to sampling n leaves from the continuous tree. It turns out that the corresponding skeleton (see the figure) of these n leaves is a random binary tree with n leaves, uniformly distributed over all binary trees with n leaves. See Aldous [1, 2, 3] for everything one may want to know about Brownian trees.

Sampling more and more points on a given Brownian excursion B , one gets a growing family of binary trees with the same law as the growing family obtained by Rémy's algorithm. Then the corresponding family of random walks converges almost surely to a Brownian excursion, as shown in Section 4, and it is easily seen that indeed, it converges uniformly to B .

3.5.2 Urns

Consider an urn with balls of two colours where sequentially, balls are picked at random and then replaced into the urn together with some extra balls. Say that this urn has replacement rule $(a, b; c, d)$ and initial condition (e, f) if

- There are initially e blue balls and f red balls
- If the ball picked at random in the urn is blue, it is replaced into the urn together with a more blue balls and b more red balls.

- If the ball picked at random in the urn is red, it is replaced into the urn together with c more blue balls and d more red ball.

It is easily seen that such urns model can be embedded in our construction. Indeed, recall the definition of L_n in Section 4. Then it follows from the analysis in Section 4.2 that $(2n, L_n)$ has the same law as the number of blue, resp. red balls in the urn. In particular, the limit laws are naturally given by the corresponding laws of the limiting Brownian paths. This leads to

Corollaire 1 *Consider an urn with replacement rule $(2, 0; 1, 1)$. Let R_n be the number of red balls after n steps, and set $r_n = R_n/\sqrt{n}$. Then*

(i) *If the initial condition is $(0, 1)$, then r_n converges almost surely to a random variable with density $c_1 x e^{-c_2 x^2}$, $x > 0$, i.e. following the Rayleigh law.*

(ii) *If the initial condition is $(1, 1)$, then r_n converges almost surely to a random variable with density $c_3 e^{-c_4 x^2}$, $x > 0$, i.e. with the law of the absolute value of a normal random variable.*

For (i), this corresponds to the height of a point chosen at random in a Brownian excursion. For (ii), this corresponds to $|B_1|$ where B is a Brownian motion. Corollary 1 can also be derived from an analysis of the Chinese restaurant ([15], Chapter 3.3) or by methods of analytic combinatorics [8]. See [10] for classical results on urns, or [8] for more recent results.

3.5.3 Random partitions

Let us recall some facts on random partitions in this section (see [15] for details). A random partition of a finite or countable state space is exchangeable if its law is invariant by permutations. If a partition is exchangeable, it is a classical consequence of De Finetti's theorem that each block B of the partition almost surely has an asymptotic frequency, which is the limit of $|B \cap \{1, 2, \dots, n\}|/n$. An interesting means of constructing random exchangeable partitions on the set of integers is the so-called Chinese restaurant introduced by Pitman, and which is a kind of urn model with infinitely many colours.

Imagine a restaurant with infinitely many tables, labeled by the integers. The first customer arrives and sits at the first table. The second customer arrives and sits either at the first table with probability $(1 - \alpha)/(1 + \theta)$ or at the second table with probability $(\alpha + \theta)/(1 + \theta)$.

More generally suppose that at a given moment, n customers have arrived and occupy k tables, the number of customers at each table being $n_1 \dots n_k$ respectively with $n_1 + \dots + n_k = n$. Then the $(n + 1)$ -th customer sits at table number i , $1 \leq i \leq k$, with probability $(n_i - \alpha)/(n + \theta)$, and at table number $k + 1$ with probability $(k\alpha + \theta)/(n + \theta)$.

Associate with this process a partition of \mathbb{N} by saying that i and j are in the same block of the partition if and only if the i -th and j -th customers sit at the same table. Then one can check that this random partition is exchangeable.

The set of return times to 0 of a simple random walk also generates a random partition \mathcal{P} . Indeed, consider a random walk of length $2n$ and say that two integers $1 \leq i, j \leq n$ are in the same block of \mathcal{P} if there is no zero of the random walk between times $2i - 1$ and $2j - 1$.

It is clear that \mathcal{P} is not exchangeable. However, one can construct an exchangeable partition \mathcal{P}' by choosing a random, uniform permutation σ on $\{1, 2, \dots, n\}$ and saying that i and j are in the same block of \mathcal{P}' if and only if $\sigma(i)$ and $\sigma(j)$ are in the same block of \mathcal{P} .

A link between the Chinese restaurant and partitions obtained from a random walk is the following result, which was first proved by Pitman and Yor [16] :

Theorem 4 (Pitman-Yor) (i) *The exchangeable partition obtained from a simple random walk of length $2n$ has the same law as the partition obtained from the first n customers of a Chinese restaurant with parameters $(1/2, 0)$.*

(ii) *The exchangeable partition obtained from a simple random walk of length $2n$ conditioned to return to 0 at time $2n$ has the same law as the partition obtained from the first n customers of a Chinese restaurant with parameters $(1/2, 1/2)$.*

This result has a continuous counterpart in terms of partitions derived from a Brownian motion or from a Brownian bridge [16].

Theorem 4 can be viewed as a direct consequence of the algorithms of Theorem 1 and of Case (1) in Theorem 2. Indeed, at each iteration of the algorithm, two steps are added and either they are incorporated to an excursion or to the meander, or these two steps form a new excursion, and the respective probabilities correspond to those given by the Chinese restaurant.

For instance, for the bridge, consider an excursion of length $2m$, starting at, say l and ending at $l + 2m$. Then the algorithm adds two steps to this excursion if and only if the random time t lies in $[l + 1, l + 2m - 1]$. This happens with probability proportional to $2m - 1$ or, equivalently, with probability proportional to $m - 1/2$. On the other hand, if the bridge consists of k excursions, then there are $k + 1$ zeros. Moreover, the algorithm adds a new excursion if and only if $S^n(t) = 0$. This happens with probability proportional to $k + 1$ or, equivalently, with probability proportional to $k/2 + 1/2$. These are exactly the probabilities for the Chinese restaurant with parameters $(1/2, 1/2)$.

3.5.4 Proof of Theorem 3

To generate the pseudobridge, the algorithm is the same as for the bridge, except that the random time t is chosen with probability proportional to

- 1 if $S^n(t) > 0$
- $1 - 1/k$ if $S^n(t) = 0$, where k denotes the number of zeros of S^n .

It should be clear that one obtains this way paths with the distribution described in Theorem 3, and that one has almost sure convergence, by the same arguments as in Section 4.

What remains to show is that the limiting object \widehat{B} is a Brownian pseudobridge. In order to do so, we prove the following properties of \widehat{B} .

- \widehat{B} is a concatenation of Brownian excursions with total length 1. Let Exc denote the set of these excursions and $l(e)$ denote the length of an excursion e .
- The family $(l(e), e \in Exc)$ has the same law as the family of asymptotic frequencies of a Chinese restaurant with parameters $(1/2, 0)$.
- Conditionally on $(l(e), e \in Exc)$, the excursions are independent,
- For every n , conditionally on Exc , let e_1, \dots, e_n be excursions of Exc chosen independently with probability proportional to the length. Then conditionally on the fact that e_1, \dots, e_n are all different, the order in which e_1, \dots, e_n occur in \widehat{B} is uniformly distributed over all possible orders.

Indeed, it is easily seen that by the same argument as in the previous subsection, one can embed the Chinese restaurant with parameters $(1/2, 0)$ in the construction given by the algorithm of Theorem 3. Moreover, by the same arguments as in the proof of (2) in Theorem 2, each excursion e of the random walk constructed by the algorithm converges almost surely to a Brownian excursion. The independence of the excursions, conditionally on their lengths, is also clear from the definition of the algorithm. Finally, the fourth point is easily seen by induction on n .

It remains to remark that a normalized pseudobridge \widehat{B}' can be constructed as follows [4]. Let S be a stable subordinator S with index $1/2$. Associate with each jump (S_{t-}, S_t) , $0 \leq t \leq 1$, an independent Brownian excursion e_t with length $(S_t - S_{t-})/S_1$. Then \widehat{B}' is the concatenation of the excursions e_t .

It turns out [16] that the family $(S_t - S_{t-})/S_1$, $0 \leq t \leq 1$ has the same law as the family of asymptotic frequencies of a Chinese restaurant with parameters $(1/2, 0)$. Moreover, it is clear that if we choose n jumps of a subordinator with probability proportional to the length, the order in which these jumps occur is uniformly distributed over all possible orders. Thus we see that the properties listed above characterize the normalized pseudobridge, and therefore \widehat{B} is a normalized pseudobridge.

Acknowledgements My interest on this subject was raised by Jim Pitman's course in St-Flour. I thank Jean-François Marckert and Philippe Duchon for references.

Bibliographie

- [1] Aldous, David. The continuum random tree. I. *Ann. Probab.* **19** (1991), no. 1, 1–28.
- [2] Aldous, David. The continuum random tree. II. An overview. *Stochastic analysis (Durham, 1990)*, 23–70, London Math. Soc. Lecture Note Ser., 167, Cambridge Univ. Press, Cambridge, 1991
- [3] Aldous, David. The continuum random tree. III. *Ann. Probab.* **21** (1993), no. 1, 248–289.
- [4] Biane, Ph. ; Le Gall, J.-F. ; Yor, M. Un processus qui ressemble au pont brownien. *Séminaire de Probabilités, XXI*, 270–275, Lecture Notes in Math., 1247, Springer, Berlin, 1987.
- [5] Csörgő, Sándor ; Hall, Peter. The Komlós-Major-Tusnády approximations and their applications. *Austral. J. Statist.* **26** (1984), no. 2, 189–218.
- [6] Csörgő, M. ; Révész, P. A new method to prove Strassen type laws of invariance principle. I, II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31** (1974/75), 255–259 ; *ibid.* **31** (1974/75), 261–269.
- [7] Donsker, Monroe D. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.* (1951). no. 6.
- [8] Flajolet, Gabarró and Pekari. Anaytic urns *Ann. Probab.* **33** (2005), no. 3, 1200–1233.
- [9] Le Gall, Jean-François ; Le Jan, Yves. Branching processes in Lévy processes : the exploration process. *Ann. Probab.* **26** (1998), no. 1, 213–252.
- [10] Johnson, Norman L. ; Kotz, Samuel. *Urn models and their application.* Wiley, 1977.
- [11] Kiefer, J. On the deviations in the Skorokhod-Strassen approximation scheme. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **13** 1969 321–332.
- [12] Knight, Frank B. On the random walk and Brownian motion. *Trans. Amer. Math. Soc.* **103** 1962 218–228.
- [13] Komlós, J. ; Major, P. ; Tusnády, G. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** (1975), 111–131.

- [14] Komlós, J. ; Major, P. ; Tusnády, G. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **34** (1976), no. 1, 33–58.
- [15] Pitman, Jim. *Combinatorial stochastic processes* (Saint-Flour,2002). To appear, Lecture Notes in Math., Springer.
- [16] Pitman, Jim ; Yor, Marc. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25** (1997), no. 2, 855–900.
- [17] Rémy, Jean-Luc. Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire. *RAIRO Inform. Théor.* **19** (1985), no. 2, 179–195.
- [18] Revuz, Daniel ; Yor, Marc. *Continuous martingales and Brownian motion*. Third edition. Springer-Verlag, Berlin, 1999.
- [19] Strassen, Volker. Almost sure behavior of sums of independent random variables and martingales. 1967 *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability* (Berkeley, Calif., 1965/66) Vol. II : Contributions to Probability Theory, Part 1, pp. 315–343.

Chapitre 4

Erratum for the paper ”Nested regenerative sets and their associated fragmentation process”

There was a mistake in the computation of the coefficients in Theorem 4.1 of the aforementioned paper. Here is a correct statement :

Théorème 1 *The process $(\mathcal{P}_\alpha, 0 \leq 1 \leq \alpha)$, is an inhomogeneous fragmentation. Its dislocation measure is given by*

$$\mu_\alpha = \frac{1}{1-\alpha} \sum_{n=1}^{\infty} c_n(\alpha) M(\alpha, -\alpha; n, 1)$$

where the coefficients $c_n(\alpha)$ are given by $c_1(\alpha) = 1$ and by the induction formula :

$$c_{n+1}(\alpha) = \frac{n-\alpha}{n+1-\alpha} c_n(\alpha).$$

The error lies in Proposition 4.1, which should be replaced by the following :

Proposition 1 *Assume that there is a dislocation at time α . Then for all integers k, n with $n \geq k$,*

$$\begin{aligned} & \mathbb{P}(H_k(\alpha-) = n+1 | I(\alpha) = k) \\ &= \frac{n!}{(1-\alpha)(2-\alpha)\dots(n-\alpha)} [x^n] \frac{[(1-\alpha)\log(1/(1-x))]^{k-1}}{(k-1)!} \end{aligned}$$

where we use the standard notation : if $f(x)$ is a function that is analytic in x , we denote by $[x^n] f(x)$ the coefficient of x^n in the expansion of $f(x)$ as a power series.

In order to prove Proposition 1, we state two lemmas.

First, for a given $\alpha \in [0, 1]$ we devise a seating plan on the on the points in $B(\alpha)$, grouping these points into subtables :

- For every $k \geq 1$, say that the point $(x_{H_k(\alpha-)}(\alpha-), z_{H_k(\alpha-)}(\alpha-))$ sits at the k -th subtable.
- For any other point, say $(x, z) \in B(\alpha)$, let

$$N(x, z) = \max\{n, z_{H_n(\alpha-)}(\alpha-) > z\}.$$

Then we say that (x, z) sits at the k -th subtable if

$$x_{H_k(\alpha-)}(\alpha-) = \max\{x_{H_n(\alpha-)}(\alpha-) \text{ s.t. } x_{H_n(\alpha-)}(\alpha-) < x \text{ and } n \leq N(x, z)\}$$

Lemma 1 *The seating plan described above yields a Chinese restaurant with parameters $(0, 1 - \alpha)$ and $k_0 = 1, a_1 = 1$.*

The proof of Lemma 1 is the same as the proof of the seating plan in Theorem 3.1 of the paper.

Second, we have :

Lemma 2 *Consider a Chinese restaurant with parameters $(0, 1 - \alpha)$ and $k_0 = 1, a_1 = 1$. For $k \geq 2$ let τ_k be the first time at which a customer sits at the k -th subtable. Then*

$$\mathbb{P}(\tau_k = n + 1) = \frac{D(k, n)(1 - \alpha)^{k-1}}{(1 - \alpha)(2 - \alpha) \dots (n - \alpha)}$$

where the coefficients $D(k, n)$ are given by

$$\sum_{n \geq 1} |D(k, n)| \frac{t^n}{n!} = \frac{(-\log(1 - t))^{k-1}}{(k - 1)!}$$

Proof of Lemma 2

First, if \prec is a total order on $\{1, 2, \dots, n + 1\}$, set $m_1 = 1$ and by induction, $m_{i+1} = \min\{j > m_i, j \prec m_i\}$. Let $D(k, n)$ be set of total orders \prec on $\{1, 2, \dots, n + 1\}$ such that $m_k = n + 1$. As an elementary application of the theory of combinatorial species,

$$\sum_{n \geq 1} |D(k, n)| \frac{t^n}{n!} = \frac{(-\log(1 - t))^{k-1}}{(k - 1)!}$$

Next, we can generate the Chinese restaurant using the points in $B(\alpha)$ and the seating rule described above. In that framework $\tau_k = n + 1$ if and only if, for the first $n + 1$ points (by decreasing y -coordinate) of $B(\alpha)$, the restriction

of the order \prec_α lies in $D(k, n)$. Moreover, it is easily seen that each of these configurations has probability

$$\frac{(1 - \alpha)^{k-1}}{(1 - \alpha)(2 - \alpha) \dots (n - \alpha)}$$

Hence the result. \square

Proof of Proposition 1

By Lemma 1, the probability

$$\mathbb{P}(H_k(\alpha-) = n + 1 | I(\alpha) = k)$$

is exactly the probability that $\tau_k = n + 1$ for a Chinese restaurant with parameters $(0, 1 - \alpha)$ and $k_0 = 1, a_1 = 1$. Lemma 2 then gives the result. \square

Proof of Theorem 1

We use Propositions 4.2 and 4.3 of the paper, together with Proposition 1. This implies that the coefficients $c_n(\alpha)$ are given by

$$\begin{aligned} c_n(\alpha) &= \sum_{k=1}^n \mathbb{P}(H_k(\alpha-) = n + 1 | I(\alpha) = k) \\ &= \frac{n!}{(1 - \alpha)(2 - \alpha) \dots (n - \alpha)} [x^n] \sum_{k=1}^n \frac{[(1 - \alpha) \log(1/(1 - x))]^{k-1}}{(k - 1)!} \\ &= \frac{n!}{(1 - \alpha)(2 - \alpha) \dots (n - \alpha)} [x^n] \exp[(1 - \alpha) \log(1/(1 - x))] \\ &= \frac{n!}{(1 - \alpha)(2 - \alpha) \dots (n - \alpha)} [x^n] \left(\frac{1}{1 - x} \right)^{1 - \alpha} \end{aligned}$$

Developing $(1 - x)^{\alpha-1}$ as a power series yields Theorem 1.