

Université de Versailles-Saint-Quentin en Yvelines

Habilitation à Diriger des Recherches

Spécialité : Mathématiques

THÈME : STRUCTURES ARBORESCENTES, ALGORITHMES

ET

THÉORÈMES LIMITES

Jean-François Marckert

Rapporteurs

| | |
|------------------------|---------------|
| Jean- François Le Gall | Paris VI -ENS |
| Jim Pitman | Berkeley |

Membres du jury

| | |
|------------------------|-------------------|
| Philippe Chassaing | Nancy I |
| Luc Devroye | Montréal (McGill) |
| Jean- François Le Gall | Paris VI -ENS |
| Didier Piau | Lyon I |
| Alain Rouault | Versailles |
| Wendelin Werner | Paris XI |

Soutenue le 14 décembre 2004

Remerciements...

Je voudrais tout d'abord remercier mes coauteurs A. Akhavi, P. Chassaing, B. Chauvin, M. Drmota, T. Klein, S. Janson, A. Makkadem, A. Panholzer, A. Rouault et M. Yor. Vous m'avez beaucoup appris et apporté. Ce fût un réel plaisir de travailler avec vous. En particulier, merci à Mokka, avec qui une bonne partie des travaux post-thèse a été réalisée, pour ces heures de recherches passionnées...

Je dois dire que j'ai eu beaucoup de chance lors de la constitution de mon jury. Mes juges, découvreurs (ou auteurs?) de puissantes et surtout jolies mathématiques, m'ont tous directement inspiré lors de l'élaboration de ces travaux. Je suis particulièrement fier qu'ils aient accepté – sans se faire prier le moins du monde – de faire partie de mon jury.

Je ne veux pas trop insister sur les relations d'amitiés qui me lient avec certains membres de peur de discréditer mon travail... Merci à vous, Philippe, Jean-François, Luc, Didier, Alain et Wendelin d'avoir accepté de vous déplacer spécialement pour ma soutenance.

Et aussi, un grand merci à Jean-François et Jim pour avoir très gentiment accepté la lourde tâche de rapporteur.

J'ai bénéficié à Versailles, au laboratoire de mathématiques, d'un environnement particulièrement sympathique. Je salue ici mes amis de l'équipe de proba : Alain, Brigitte, Emmanuel, Julien, Mariane, Mokka, Nicolas, Oleksiy, Thierry... et tous les autres qui supportent tant bien que mal ma présence au labo, les thésards, JF, Philippe, Nathalie, Tahar, Aurélie, Stéphane, etc...

En guise de fin paradoxale, merci enfin à tous ceux qui ne sont pas remerciés.

Table des matières

| | |
|--|-----------|
| — Introduction — | 7 |
| I. Arbres, marches branchantes et quadrangulations | 11 |
| 1 Des arbres | 11 |
| 1.1 Le formalisme de Neveu | 11 |
| 1.2 Codage des arbres finis | 12 |
| 1.3 Modèle probabiliste : les arbres de Galton-Watson conditionnés par la taille | 15 |
| 1.4 Modèle et résultats asymptotiques | 18 |
| 1.5 Arbre abstrait | 25 |
| 2 Des marches branchantes | 29 |
| 2.1 Premiers travaux | 29 |
| 2.1.1 Serpents discrets et continus | 30 |
| 2.1.2 Un homéomorphisme | 33 |
| 2.2 Marche branchante : cas général | 35 |
| 2.2.1 Convergence du tour | 35 |
| 2.2.2 Convergence des serpents | 38 |
| 2.3 Une propriété asymptotique de la correspondance par rotation | 40 |
| 3 Des quadrangulations | 43 |
| 3.1 Modèles et problématique | 43 |
| 3.2 Codage des quadrangulations | 45 |
| 3.3 Notion de carte “abstraite” | 49 |
| 3.4 Convergence de (r_n, v_n) sous \mathbb{P}_D^n | 50 |
| II. Autres structures aléatoires et algorithmes | 55 |
| 1 Parking avec densité | 55 |
| 1.1 Données du problème | 55 |
| 1.2 Résultats connus concernant le parking uniforme | 56 |
| 1.3 Parking avec densité : résultats | 57 |
| 2 Éléments sur la réduction de Lovász d’un réseau dans \mathbb{R}^n | 60 |
| 2.1 Introduction | 60 |
| 2.2 Résultats | 60 |
| 3 Arbres binaires de recherche et martingales | 63 |
| 3.1 Arbres binaires de recherche et arbres de Yule | 63 |
| 3.2 Liens entre les martingales | 66 |
| 3.3 Convergence du profil | 69 |

| | |
|---|-----------|
| III. Résultats issus de la thèse | 73 |
| 1 Couple hauteur-largeur des arbres binaires | 73 |
| 2 Largeur des arbres de Cayley et parkings | 74 |
| 3 Recherche du maximum d'une marche aléatoire | 75 |
| — Références — | 76 |

INTRODUCTION

Le présent mémoire constitue un exposé succinct des travaux de recherche auxquels j'ai collaboré. J'ai eu beaucoup de mal à régler le niveau d'exposition des résultats et à décider ce que j'allais expliciter. J'ai finalement choisi de définir tous les objets mathématiques évoqués et de retranscrire la plupart des résultats ; les preuves quant à elles sont parfois esquissées, mais seulement dans des cas présentant un intérêt particulier.

Pour de plus amples détails, on pourra se tourner vers les articles dont la liste est donnée plus bas et que l'on pourra trouver sur ma page ouèbe : www.math.uvsq.fr/~marckert.

Ce mémoire est composé de trois parties dont voici une description très sommaire :

- La première partie, la plus importante de ce document, est consacrée aux arbres de Galton–Watson conditionnés par la taille et à d'autres objets associés de manière plus ou moins naturelle : les marches branchantes conditionnées par la taille et les quadrangulations de taille fixée. Nous étudions principalement le comportement asymptotique de ces objets, lorsque la taille tend vers $+\infty$. Dans chacun de ces cas, nous montrons la convergence faible des structures discrètes normalisées vers un objet continu aléatoire. Les limites respectives sont :
 - l'arbre continu d'Aldous (on donne une nouvelle preuve),
 - le serpent brownien (ou sauteur) de processus de durée de vie l'excursion brownienne,
 - la carte brownienne, que l'on définit.
- Dans la deuxième partie, je résume trois articles, indépendants, concernant des structures aléatoires issues de problèmes informatiques. Le but de ces articles est de mieux comprendre le fonctionnement de certains algorithmes. Il s'agit :
 - du problème de parking non uniforme (ou hachage linéaire non uniforme),
 - d'un problème de réduction de réseau euclidien,
 - de l'analyse de l'arbre binaire de recherche.
- Enfin, dans une troisième partie, je donne un rapide aperçu de trois articles produits à l'issue ou pendant ma thèse de doctorat.

TRAVAUX

- [1] J.F. Marckert, (2001). *Parking with density*, Random Struct. and Algorithms, 18, 4, 364-380.
- [2] J.F. Marckert & A. Panholzer, (2002). *Noncrossing trees are almost conditioned Galton-Watson trees*, Random Struct. Algorithms 20, No.1, 115-125.
- [3] J.F. Marckert & A. Mokkaïdem, (2003). *States spaces of the snake and its tour — convergence of the discrete snake*, J. Theor. Probab. 16, No.4, 1015-1046.
- [4] J.F. Marckert & A. Mokkaïdem, (2003). *Ladder variables, internal structure of Galton-Watson trees and finite branching random walks*, J. Appl. Probab. 40, No. 3, 671–689.
- [5] J.F. Marckert & A. Mokkaïdem, (2003). *The depth first processes of Galton-Watson trees converge to the same Brownian excursion*, Ann. Probab. 31, No. 3, 1655–1678.

- [6] J.F. Marckert, (2004). *The rotation correspondence is asymptotically a dilatation*, Random Struct. Algorithms 24, No.2, 118-132.
- [7] M. Drmota & J.F. Marckert, (2004). *Reinforced weak convergence of stochastic processes*, Statistics Probab. Letters, à paraître.

———— ARTICLES SOUMIS ————

- [8] B. Chauvin, T. Klein, J.F. Marckert & A. Rouault, *Martingales and Profile of Binary Search Trees*
- [9] S. Janson & J.F. Marckert, *Convergence of discrete snakes*
- [10] J.F. Marckert & A. Mokkadem, *Limit of Normalized Quadrangulations : the Brownian map*
- [11] A. Akhavi, J.F. Marckert & A. Rouault, *On the Lovász reduction of a random basis*

———— ARTICLES ISSUS DE LA THÈSE ————

- [12] P. Chassaing & J.F. Marckert, (2001). *Parking functions, empirical processes and the width of rooted labeled trees*, The Electronic Journal of Combinatorics 8, R14.
- [13] P. Chassaing, J.F. Marckert & M. Yor, (2000). *The height and width of simple trees*, in Mathematics and Computer Science : Algorithms, Trees, Combinatorics and Probabilities. D. Gardy, A. Mokkadem Ed. - Birkhäuser
- [14] P. Chassaing, J.F. Marckert & M. Yor, (2003). *A stochastically quasi-optimal search algorithm for the maximum of the simple random walk*, Ann. Appl. Probab. 13, No.4, 1264-1295.

Le reste de la bibliographie est placée en fin de mémoire.

I. Arbres, marches branchantes et quadrangulations

--*--*--*--*--*--

1 Des arbres

Dans cette première section sont exposés des résultats concernant les arbres de Galton–Watson conditionnés par la taille. Je renvoie au récent cours de Pitman à Saint–Flour [97] (principalement chap. 5 & 6) pour un état de l’art.

1.1 Le formalisme de Neveu

Nous reprenons la définition des arbres de Neveu [91] : soit $\mathbb{U} = \emptyset \cup \bigcup_{n \geq 1} \mathbb{N}^{*n}$, l’ensemble des mots finis ou vide sur l’alphabet $\{1, 2, 3, \dots\}$. On note uv la concaténation du mot u avec le mot v . Une partie E de \mathbb{U} est appelée *arbre* si elle possède les propriétés suivantes :

$$\begin{cases} \emptyset \in E \\ \text{si } uv \in E \text{ alors } u \in E \\ \text{si } uk \in E \text{ avec } k > 1 \text{ alors } u(k-1) \in E. \end{cases}$$

Les éléments de E sont appelés *nœuds*, \emptyset la *racine*. Soit $u \in E$. Les préfixes stricts de u sont appelés

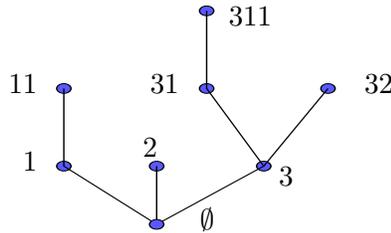


Figure 1: Représentation de l’arbre $E = \{\emptyset, 1, 2, 3, 11, 31, 32, 311\}$

ancêtres de u ; les éléments de E dont u est un préfixe strict sont appelés *descendants* de u . Les éléments de E de la forme uk (pour k dans \mathbb{N}^*) sont appelés *fil*s de u . Un nœud sans descendant est appelé *feuille*, les autres *nœuds internes*. Le cardinal de E est appelé *taille de l’arbre* ou *population totale*.

Les trois conditions imposées sur les arbres imposent l’existence d’une racine, l’existence d’un chemin liant tout nœud à la racine (un arbre est connexe) et enfin, que deux arbres différents ont des structures branchantes différentes.

L’ensemble \mathbb{U} , muni de l’ordre lexicographique est un espace totalement ordonné (en prenant $\emptyset < u$ pour tout $u \in \mathbb{U} \setminus \{\emptyset\}$). Ceci induit un ordre sur les nœuds de chaque arbre.

On représentera souvent les arbres comme sur la Figure 1 en omettant les étiquettes ; il est alors sous entendu que la racine est en bas, que les fils de chaque nœud du plus petit au plus grand sont représentés du plus à gauche au plus à droite. Ceci permet d’établir un lien avec la méthode combinatoire où les arbres sont (parfois) considérés comme étant des *cartes enracinées à une seule face* (voir Fig. 2). Sous ce point de vue, un représentant est un graphe connexe sans cycle plongé dans le plan, dont les nœuds ne possèdent pas d’étiquette et dont une arête orientée est distinguée (cette arête part du nœud racine et va au nœud 1). Un arbre est défini comme étant une classe de représentants à homéomorphisme direct du plan près. C’est ce point de vue qui sera retenu pour l’étude des quadrangulations.

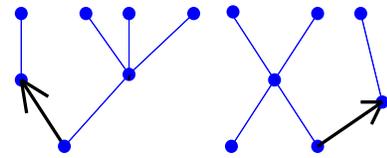


Figure 2: Deux représentations du même arbre “à la combinatoire”

1.2 Codage des arbres finis

Une méthode particulièrement efficace pour étudier les arbres consiste à les coder par des chemins. Pour ce faire, on utilise des algorithmes de *parcours* basés sur le système de numérotation lexicographique (voir Knuth [77] et Flajolet & Sedgewick [61] pour d'autres parcours).

PARCOURS EN LARGEUR. On visite successivement les nœuds de l'arbre comme suit : d'abord on visite la racine, puis les nœuds d'une lettre par ordre lexicographique, puis les nœuds de deux lettres par ordre lexicographique, etc...

PARCOURS EN PROFONDEUR. On visite les nœuds selon l'ordre lexicographique.

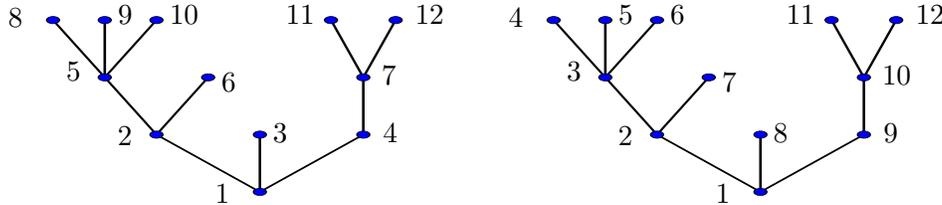


Figure 3: Ordre de parcours en largeur et en profondeur

Nous utilisons les algorithmes de parcours afin de *coder* les arbres par des chemins.

Représentation des arbres par des chemins

Soit E un arbre de cardinal fini. On note Y_i le nombre de fils du i -ème nœud visité (Y_1 =nombre de fils de la racine) lors du parcours en profondeur de E (par exemple, la forme de l'arbre de la Fig. 3 est $(3, 2, 3, 0, 0, 0, 0, 1, 2, 0, 0)$). Il est facile de vérifier que la liste $(Y_i)_{i=1, \dots, \#E}$ détermine complètement E . On appellera cette liste la *forme* de l'arbre.

Soit $n = \#E$. On appelle *pile du parcours en profondeur* (PPP) associée à E , la marche $(S_j)_{j=0, \dots, n}$ définie par :

$$S_0 = 0, \quad \text{et pour } 1 \leq j \leq n, \quad S_j = (Y_1 - 1) + \dots + (Y_j - 1).$$

Il est facile de voir que la PPP caractérise E .

CONTRAINDRE LA TAILLE

Il apparaît qu'une liste de nombres entiers (Y_1, \dots, Y_n) est une forme d'arbre de taille n sssi elle satisfait à l'ensemble de contraintes suivantes :

$$(C) = \begin{cases} \sum_{i=1}^k Y_i \geq k & \text{pour tout } k \in \llbracket 1, n-1 \rrbracket, \\ Y_1 + \dots + Y_n = n-1. \end{cases}$$

Cette contrainte sur la liste (Y_1, \dots, Y_n) est équivalente à la contrainte suivante sur la PPP :

$$(C') = \begin{cases} S_0 = 0, \\ S_k \geq 0 & \text{pour tout } k \in \llbracket 1, n-1 \rrbracket, \\ S_n = -1. \end{cases}$$

Tous se passe de la même manière avec le parcours en largeur ; on appellera *pile du parcours en largeur* (PPL) la marche S construite avec les variables Y_i issues du parcours en largeur.

Dans la littérature, ces marches sont parfois appelées marches de Łukaciewicz.

Parcours en largeur et codage

On appelle *profil* de l'arbre T la suite $(Z_k)_{k \geq 0}$ où Z_k est le nombre de nœuds de T à distance k de la racine. La largeur W et la hauteur H de T sont définies par $W = \max\{Z_k, k \geq 0\}$ et $H = \max\{k, Z_k > 0\}$. Le profil se lit sur la PPL (voir Fig. 4), on a

$$Z_0 = 1, \quad Z_1 = 1 + S_{1+Z_0}, \quad \dots, \quad Z_k = 1 + S_{1+Z_0+\dots+Z_{k-1}}, \quad \dots$$

Cette propriété est utilisée dans les Sections 1 et 2 de la partie III.

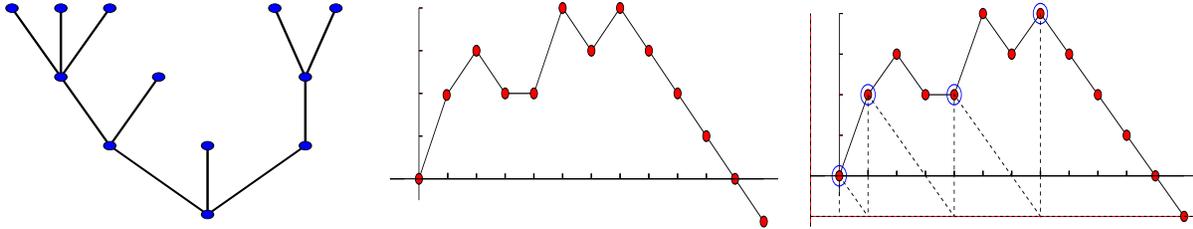


Figure 4: Un arbre, sa PPL ; sur le troisième dessin sont entourés les points où on lit le profil

Parcours en profondeur et codage

En plus de la PPP, le parcours en profondeur permet d'associer à chaque arbre deux chemins : le tour de l'arbre et le processus des hauteurs (voir aussi Fig. 5).

LE TOUR DE L'ARBRE (OU MARCHE DE HARRIS [65]) : soit T un arbre ayant n nœuds. On construit une fonction F , que l'on appellera fonction du parcours en profondeur (FPP) définie de l'ensemble $\llbracket 0, 2(n-1) \rrbracket$ à valeurs dans T . On pose $F(0) = \emptyset$, la racine de T . Les autres valeurs de F sont construites inductivement : si le nœud $F(k-1)$ a des fils non visités, c'est-à-dire n'appartenant pas à $\{F(0), \dots, F(k-2)\}$, alors $F(k)$ est le fils le plus petit parmi les fils de $F(k-1)$ non visités. Si tous les fils de $F(k-1)$ ont été visités, alors $F(k)$ est le père de $F(k-1)$.

On appelle *tour de l'arbre*, la fonction V de $\llbracket 0, 2(n-1) \rrbracket$ à valeurs dans \mathbb{N} définie par

$$V(k) = d(F(0), F(k)),$$

la distance dans l'arbre entre la racine et le nœud $F(k)$.

On peut décrire les fonctions F et V de manière intuitive : F donne la position d'une mouche marchant autour de l'arbre à la vitesse 1 dans le sens des aiguilles d'une montre et étant à la racine au temps 0. Le tour V est la fonction qui donne la hauteur de la mouche en fonction du temps.

LE PROCESSUS DES HAUTEURS : on note v_i le i -ème nœud visité lors du parcours en profondeur ($v_0 = \emptyset$) ; le nœud v_i est aussi le i -ème mot de l'arbre pour l'ordre lexicographique. Le processus des hauteurs est défini par

$$h(i) = d(v_0, v_i) \quad \text{pour } 0 \leq i \leq n-1.$$

UNE NOTION DE COIN

Lorsque la mouche marche autour de l'arbre, elle visite chaque nœud de degré sortant $k, k+1$ fois. Dans la représentation sous forme de carte, la position de la mouche dans l'arbre est plus précise que l'information donnée par la FPP. On fait appel à une notion de "coin" que l'on peut voir de

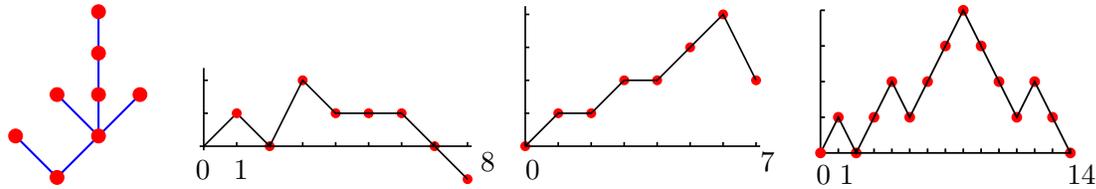


Figure 5: un arbre, sa PPP, son processus des hauteurs et son tour.

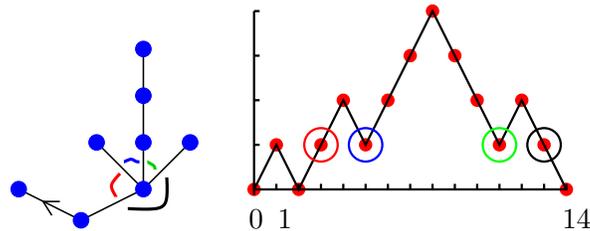


Figure 6: Représentation des 4 coins autour d'un nœud de degré 3

différentes manières : position dans l'arbre de la mouche, "secteur" entre deux arêtes adjacentes consécutives pour l'ordre cyclique, temps dans $[0, 2n]$ codant cet instant de visite,... Voir Fig. 6.

CONSTRUCTION RÉCIPROQUE

Chaque marche – la PPP, la PPL, le processus des hauteurs et le tour de l'arbre – permet de retrouver l'arbre dont elle est issue.

- En partant de la PPP ou de la PPL, on peut facilement reconstituer les mots sur l'alphabet \mathbb{U} (voir aussi Fig. 9 p.16).
- Le tour de l'arbre est très proche "géométriquement" de l'arbre : il suffit de recoller les différents représentants de chaque nœud pour obtenir l'arbre. En fait, le tour peut être vu comme une sorte d'éclatement de l'arbre.
- À partir du graphe du processus des hauteurs, le père du nœud représenté en $(k, h(k))$ est le dernier nœud ayant pour abscisse $h(k) - 1$ (voir Fig. 7).

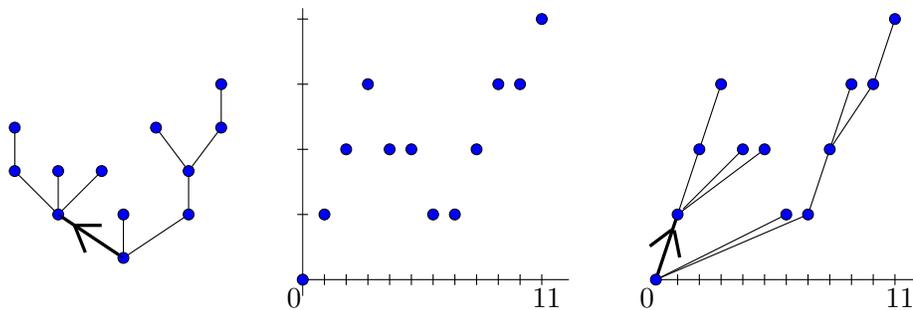


Figure 7: un arbre, son processus des hauteurs, puis reconstruction de l'arbre initial sur son processus des hauteurs

PARCOURS INVERSE ET ARBRE CROULANT

Dans la Figure 7, la représentation sous forme de carte a été choisie. Elle va permettre de mieux comprendre le paragraphe présent. Reprenons l'histoire de la mouche marchant autour de l'arbre, dans le sens des aiguilles d'une montre, commençant le long de la flèche ($\emptyset \rightarrow 1$). Le tour de l'arbre représente son ordonnée en fonction du temps et le processus des hauteurs représente les hauteurs successives des nouveaux nœuds rencontrés. Si la mouche décide de marcher dans le sens

trigonométrique en partant le long de l'arête racine (de l'autre côté de l'arête racine), le processus "ordonnée de la mouche en fonction du temps" donne un *tour inverse* et on a naturellement aussi un processus des *hauteurs inverses* h' .

■ Pour obtenir le tour inverse à partir du tour : supposons que le tour revienne $k + 1$ fois en 0. Ceci délimite k excursions. Réordonner les excursion ainsi : 1, k , $k - 1, \dots, 2$. Retourner le temps sur chaque excursion.

■ Le processus des hauteurs inverses n'est pas une fonctionnelle simple du processus des hauteurs. Pour reconstruire l'arbre en partant du processus des hauteurs inverses, on observe que le père du nœud $(k, h'(k))$ est le dernier nœud visité de hauteur $h'(k) - 1$, comme dans le processus des hauteurs. L'unique différence est que l'on visite chaque nœud après ses frères supérieurs. Pour retrouver l'arbre à partir du processus des hauteurs inverses, on doit procéder comme sur la Figure 8 pour rétablir l'ordre dans chaque fratrie. On obtient une représentation de l'arbre sous forme d'un arbre que l'on qualifie (dans [10]) de "croulant".

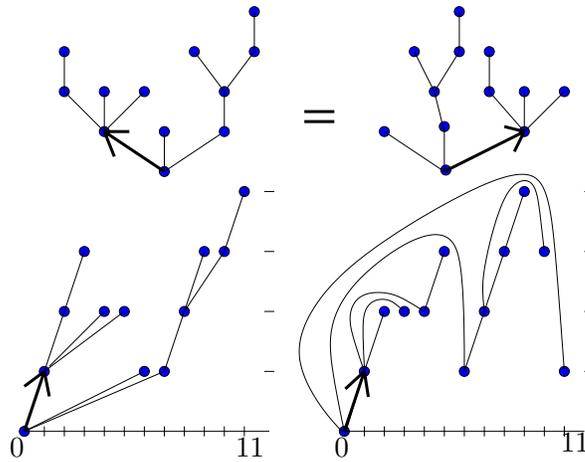


Figure 8: Au dessus, deux représentations du même arbre enraciné. À gauche, reconstruction de l'arbre grâce à son processus des hauteurs : construction d'un arbre couché. À droite, reconstruction de l'arbre grâce à son processus des hauteurs inverse : construction de l'arbre croulant

1.3 Modèle probabiliste : les arbres de Galton-Watson conditionnés par la taille

Le modèle des arbres de Galton-Watson (GW) conditionnés par la population totale est le modèle central de cette première partie. La littérature fourmille de résultats à propos des arbres de GW ; je me contenterai de rappeler le strict nécessaire afin de définir les modèles étudiés ici (on pourra se référer à Athreya & Ney [29], à Lyons & Peres [87]).

Un processus de GW est une chaîne de Markov $(Z_n)_{n=0,1,2,\dots}$ sur \mathbb{N} telle que $Z_0 = 1$ et telle que pour $k \geq 0$, Z_{k+1} s'écrit sous la forme :

$$Z_{k+1} = \sum_{l=1}^{Z_k} Y_l^{(k)}$$

où les variables $Y_l^{(k)}$ sont des copies indépendantes de loi $(p_j)_{j \geq 0}$.

Ceci modélise l'évolution d'une population se reproduisant selon les règles suivantes :

- Au temps $T = 0$, $Z_0 = 1$ individu est en vie.

- Au temps $T = k + 1$, chacun des Z_k individus vivant au temps k meurt laissant j descendants avec probabilité p_j indépendamment des autres individus.

La loi $(p_j)_{j \geq 0}$ est appelée loi de *reproduction*. L'arbre généalogique de la population $(Z_k)_{k \geq 0}$ est appelé *arbre de GW*. C'est un arbre ordonné, car on distingue les fils d'un même père. La population totale X du processus de GW est définie comme étant la taille de l'arbre généalogique :

$$X = Z_0 + Z_1 + Z_2 + \dots$$

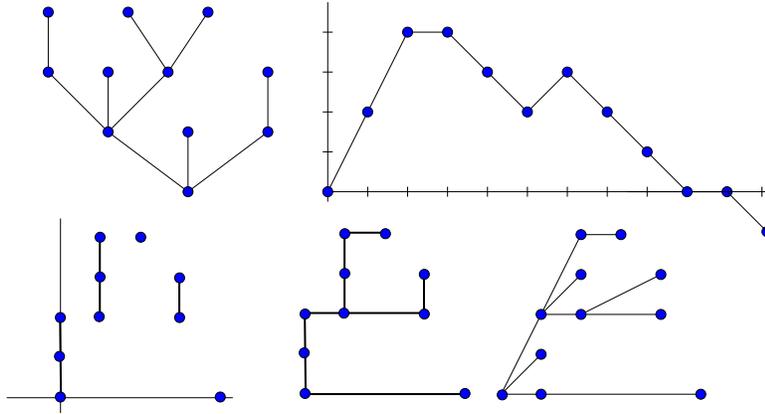


Figure 9: Sur la première ligne, un arbre τ et sa PPP. Sur le 3ème dessin, on a remplacé chaque pas $+k$ de la PPP par un bâton de hauteur $+k$, placé à la même hauteur que dans la PPP. Dans le 4ème dessin, on a construit un arbre unaire-binaire en faisant partir du point inférieur de chaque bâton (à partir du deuxième, plus exactement), vers la gauche, une ligne horizontale arrêtée au premier contact. En dernière place est l'arbre associé par la correspondance par rotation à cet arbre unaire-binaire : c'est τ ! Pour aller de l'arbre unaire-binaire à τ , on procède comme suit : on conserve les relations père-fils le plus à droite, et on remplace chaque relation père-fils gauche par la relation nœud-frère gauche. Le passage direct de la liste des bâtons à τ est également aisé.

PROBABILITÉ QU'UN ARBRE DE GW SOIT DE TAILLE n

Prenons maintenant une suite de variables Y_k de loi $(p_j)_{j \geq 0}$, indépendantes et construisons :

- un arbre en utilisant la variable Y_i comme étant le nombre de fils du i ème nœud dans le parcours en profondeur
- une marche aléatoire S , dont les incréments sont les variables $Y_k - 1$:

$$S_0 = 0 \quad \text{et} \quad S_k = \sum_{j=1}^k (Y_j - 1) \quad \text{pour } k \geq 1.$$

Il est facile de voir que l'arbre que l'on a construit est un arbre de GW de loi de progéniture $(p_j)_{j \geq 0}$, que sa PPP est S , et que sa taille est $T_{-1} = \inf\{j, S_j = -1\}$ le temps d'atteinte de -1 pour la marche S . Sa forme est donc $(Y_1, \dots, Y_{T_{-1}})$. On a ainsi $X = T_{-1}$ et donc $\mathbb{P}(X = n) = \mathbb{P}(T_{-1} = n)$. La valeur de $\mathbb{P}(T_{-1} = n)$ s'obtient par le *principe de rotation* (Dvoretzky & Motzkin [54]) :

$$\mathbb{P}(T_{-1} = n) = \frac{1}{n} \mathbb{P}(S_n = -1) \quad \text{pour tout } n \geq 1. \quad (1)$$

Si les variables $Y_i - 1$ sont **centrées** et possèdent une variance finie et non nulle σ^2 , la version locale du théorème de la limite centrale permet de donner l'équivalent suivant :

$$\mathbb{P}(X = n) \sim \frac{n^{-3/2}}{\sqrt{2\pi\sigma}}. \quad (2)$$

Ceci étant valable sous la condition supplémentaire $\text{pgcd}\{i \text{ t.q. } \mathbb{P}(Y_1 = i) > 0\} = 1$.

Puisque contraindre les variables Y_i par (C) est équivalent à contraindre la marche S par (C') , la PPP S associée à un arbre de GW conditionné par $\{X = n\}$ est simplement une marche aléatoire dont les incréments sont des copies de $Y_i - 1$, conditionnée par (C') (voir aussi Kendall [72], Takàcs [107] et Limic [83, 84]).

EFFET DU CONDITIONNEMENT PAR $\{X = n\}$ SUR LA PROBABILITÉ D'UN ARBRE

Soit (y_1, \dots, y_n) une forme d'arbre de taille n . Sachant $\{X = n\}$, quelle est la probabilité qu'un arbre de GW ait la forme (y_1, \dots, y_n) ?

Remarquons d'abord que sans le conditionnement, la réponse est simple :

$$\mathbb{P}((Y_1, \dots, Y_n) = (y_1, \dots, y_n)) = \prod_{i=1}^n p_{y_i}. \quad (3)$$

Le conditionnement ne va pas changer fondamentalement les choses, les probabilités obtenues étant proportionnelles à celles avant conditionnement :

$$\begin{aligned} \mathbb{P}((Y_1, \dots, Y_n) = (y_1, \dots, y_n) | X = n) &= \frac{\mathbb{P}((Y_1, \dots, Y_n) = (y_1, \dots, y_n))}{\mathbb{P}(X = n)} \\ &= \frac{\prod_{i=1}^n p_{y_i}}{\mathbb{P}(X = n)} \end{aligned} \quad (4)$$

EXEMPLES DE FAMILLES D'ARBRES DE GW CONDITIONNÉS PAR LA TAILLE

■ Arbres binaires : on appelle arbre binaire de GW, un arbre sous une loi de reproduction de type $p_2 = p, p_0 = 1 - p$ où $p \in]0, 1[$. Conditionnons par $\{X = 2n + 1\}$. Un arbre binaire ayant $2n + 1$ nœuds possède n nœuds internes et $n + 1$ feuilles. Ainsi, sachant $\{X = 2n + 1\}$, la probabilité d'obtenir un arbre donné de taille n est $(1 - p)^{n+1} p^n / \mathbb{P}(X = 2n + 1)$. Ceci ne dépend pas de la forme de l'arbre : donc, sachant $X = 2n + 1$, tous les arbres binaires de taille $2n + 1$ sont équiprobables.

■ Arbres planaires (ou géométriques) : on considère comme loi de reproduction la loi géométrique $p_i = (1 - p)^i p$ pour $p \in]0, 1[$ fixé. On a

$$\mathbb{P}((Y_1, \dots, Y_n) = (y_1, \dots, y_n) | X = n) = \frac{\prod_{i=1}^n ((1 - p)^{y_i} p)}{\mathbb{P}(X = n)} = \frac{p^n (1 - p)^{n-1}}{\mathbb{P}(X = n)}$$

car chaque nœud sauf la racine est le fils d'un autre nœud. La probabilité ne dépend pas de la forme de l'arbre et donc tous les arbres obtenus sont équiprobables. D'autre part, la loi géométrique attribue une probabilité non nulle à tous les arbres de taille n . Ainsi, la loi induite par le conditionnement par $\{X = n\}$ est la loi uniforme sur l'ensemble des arbres de taille n .

■ Arbres sans croisements : soit \mathcal{C} un cercle; considérons n points choisis sur \mathcal{C} , numérotés de 1 à n dans le sens des aiguilles d'une montre (comme sur la Fig. 10). On appelle arbre sans croisements (SC) un arbre ayant pour sommets ces n points, dont les arêtes sont des segments et tels que les arêtes ne se croisent pas. On suppose maintenant que la racine est placée sur le nœud 1. Pour étudier la forme des arbres SC, on efface le cercle et les numéros, on conserve les relations d'ordre ainsi que les relations père-fils (comme sur la figure, on suspend en quelque sorte l'arbre par le nœud 1 et on peigne tout vers le bas).

Avec A. Panholzer [2], nous avons montré que les arbres sans étiquettes sous-jacents étaient des arbres de GW conditionnés par $\{X = n\}$, à ceci près que la loi de reproduction de la racine est différente de celle des autres nœuds. Ceci permet de retrouver la totalité des résultats asymptotiques connus concernant les arbres SC et également de trouver des résultats nouveaux concernant la convergence de leur tour et de leur profil.

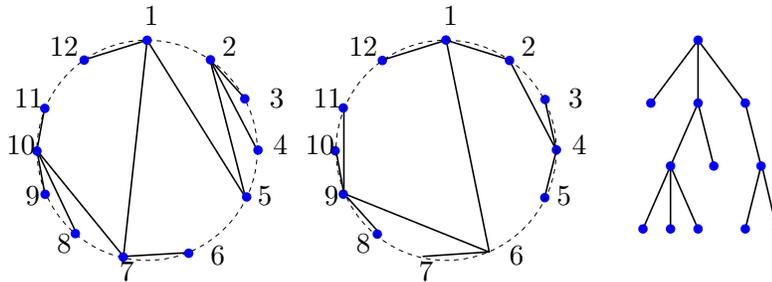


Figure 10: Deux arbres sans croisements de taille 12 et leur représentation hors cercle.

Famille exponentielle et normalisation

La moyenne de la loi de reproduction joue un rôle important dans le comportement des arbres de GW. Sous le conditionnement par $\{X = n\}$, on a vu sur les exemples que ce n'était plus le cas.

Soit $(p_k)_{k \geq 0}$ une loi de reproduction de fonction génératrice f , de rayon de convergence R . Soit $a \in]0, R[$. On définit la loi de reproduction $(\tilde{p}_k)_{k \geq 0}$ par

$$\tilde{p}_k = \frac{a^k p_k}{f(a)};$$

les lois p et \tilde{p} seront dites de la même famille exponentielle (terminologie de Kennedy [73], dont est tiré ce paragraphe). Notons avec des tildes les variables sous la loi \tilde{p} . On a

$$\mathbb{P}(\tilde{Y}_1 = y_1, \dots, \tilde{Y}_n = y_n \mid \tilde{X} = n) = \frac{\prod_{k=1}^n a^{y_k} p_{y_k}}{f(a)^n \mathbb{P}(\tilde{X} = n)} = \frac{\prod_{k=1}^n p_{y_k}}{\mathbb{P}(\tilde{X} = n) f(a)^n a^{-n+1}}$$

ce qui est proportionnel à

$$P(Y_1 = y_1, \dots, Y_n = y_n \mid X = n) = \frac{\prod_{k=1}^n p_{y_k}}{\mathbb{P}(X = n)}.$$

Ces deux probabilités conditionnelles sont égales puisque proportionnelles. Ainsi, la loi induite par p et \tilde{p} sous le conditionnement par $\{X = n\}$ est la même loi de probabilité.

La valeur de a , lorsqu'elle existe, telle que $af'(a) = f(a)$ fournit le cas *critique*, c'est-à-dire le cas où $\tilde{m} = 1$. La variance $\tilde{\sigma}^2$ de la loi de reproduction sous cette loi, lorsqu'elle est finie est égale à $a^2 f''(a)/f(a)$. Cette variance va jouer dans la suite un rôle normalisateur de tout premier ordre puisque l'on va travailler à partir de maintenant sous le modèle critique.

Commentaires

Le lien entre les modèles d'arbres dits "simples" de la combinatoire et les arbres de GW conditionnés est donné dans Aldous [19]. On trouve dans la littérature d'autres conditionnements que par la taille : par exemple, conditionnement par la hauteur [62], conditionnement par la non-extinction [87, 88].

1.4 Modèle et résultats asymptotiques

Dans la suite, on supposera toujours que la loi de reproduction $(p_j)_{j \geq 0}$ est critique (c'est-à-dire de moyenne 1), de variance $\sigma^2 \in]0, +\infty[$ et possède des moments exponentiels dans un voisinage de 0. Nous noterons \mathbb{P}_n la loi conditionnelle sachant $\{X = n\}$.

Dans cette partie nous nous intéressons à l'asymptotique des arbres de GW sous la loi \mathbb{P}_n lorsque n tend vers $+\infty$. Nous obtenons deux sortes de résultats :

- Le premier type de résultats concerne les limites en loi de certaines caractéristiques de l'arbre.
- Le deuxième type de résultats concerne la convergence globale de l'arbre vers un arbre continu.

On note h_n , V_n et S_n respectivement le processus des hauteurs, le tour et la PPP associés à un arbre de GW sous la loi \mathbb{P}_n . À partir de maintenant, on considère que ces processus sont interpolés linéairement par morceaux entre les points de coordonnées entières et sont donc continus.

Le premier résultat que nous mentionnons est le suivant : les trois processus h_n, V_n, S_n sont asymptotiquement proportionnels :

Théorème 1 [5] *La convergence faible suivante a lieu dans $C[0, 1]^3$*

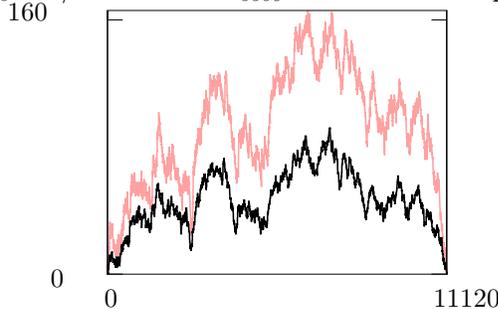
$$\left(\frac{h_n(n.)}{\sqrt{n}}, \frac{V_n(2n.)}{\sqrt{n}}, \frac{S_n(n.)}{\sqrt{n}} \right) \xrightarrow[n]{\text{faible}} \left(\frac{2}{\sigma}e, \frac{2}{\sigma}e, \sigma e \right)$$

où $(e_s)_{s \in [0,1]}$ est l'excursion brownienne normalisée.

Commentaires

- Il faut absolument remarquer qu'il s'agit de la même excursion brownienne dans le membre de droite! Ainsi on a par exemple que $(\frac{\sigma^2}{2} \frac{h_n(nt)}{\sqrt{n}} - \frac{S_n(nt)}{\sqrt{n}})_{t \in [0,1]}$ tend faiblement vers le processus p.s. nul sur $[0, 1]$. On voit très bien ce phénomène sur les simulations dès que la taille de l'arbre dépasse quelques dizaines et que la loi de reproduction, critique, possède des moments assez grands (les moments exponentiels ne sont pas nécessaires).

Sur la figure suivante, on a représenté en gris la PPP et en noir le tour d'un arbre de GW de distribution $p_0 = 13/18, p_2 = 1/6, p_6 = 1/9$ sous la loi \mathbb{P}_{5560} . La valeur correspondante de $\sigma^2/2$ est $11/6$.



La quasi-proportionnalité des deux courbes se devine.

- La convergence de la marginale “tour” était connue et due à Aldous [18, 19, 20], la convergence de la PPP due à Kaigh [70], et celle du processus des hauteurs n'était pas difficile à prouver en utilisant la convergence du tour (on va y revenir). D'autre part, une approche directe de la convergence du tour est donnée dans Bennies & Kersting [30]. Cette approche a pour point commun avec la nôtre qu'elle explique le rapport de $\sigma^2/2$ entre le tour et la Lukaciewicz en utilisant des variables d'échelles.
- Dans l'article original [5], on montrait que deux autres processus convergeaient, à un facteur multiplicatif près vers la même excursion brownienne : il s'agissait du processus des profondeurs des nœuds de degré k (lors du parcours de l'arbre dans l'ordre lexicographique) et du processus des profondeurs des nœuds racines d'un sous arbre fixé. Ceci témoigne d'une grande uniformité dans l'arbre.
- Un grand avantage de ce théorème est sa preuve, qui est vraiment très simple (elle est esquissée plus loin). On prouve directement la convergence du tour de l'arbre V_n vers l'excursion brownienne, sans passer par l'existence d'un arbre limite, comme le faisait originalement Aldous. Le bénéfice est la construction élémentaire d'une notion d'arbre continu aléatoire, limite des arbres de GW normalisés, conditionnés par la taille. Dans la Section 1.5, nous présentons cette construction.
- On donne une raison simple du rapport $\sigma^2/2$ entre le tour de l'arbre et la PPP (ceci est d'ailleurs notre explication du fait que $V(2n.)/\sqrt{n}$ converge vers $2e/\sigma$).
- Lorsque la loi de reproduction (sous la forme critique) ne possède pas de variance, il n'y a plus convergence vers l'excursion brownienne. Voir les œuvres de Duquesne [50], Duquesne & Le Gall [51, 52] & Le Gall & Le Jan [80, 79] à propos des arbres de Lévy (et de la convergence vers ceux-ci).

J'esquisse maintenant la preuve de ce théorème. Au passage, d'autres résultats seront aussi énoncés. Deux idées directrices nous poussent :

- ★ De tous les processus associés aux arbres, la PPP est celui dont on peut dire le plus de choses : il s'agit d'une marche aléatoire conditionnée, elle est markovienne, etc...
- ★ On peut passer de la PPP au processus des hauteurs et au tour par une transformation assez sympathique. On pourra alors tirer parti du fait que le tour est très proche topologiquement de l'arbre pour obtenir des renseignements sur icelui.

Lien entre le processus des hauteurs et le tour de l'arbre

- À partir du tour, il est aisé de construire le processus des hauteurs : il suffit de conserver les points qui suivent une montée, comme sur la première colonne de la Figure 11.
- À partir du processus des hauteurs, le tour s'obtient par une sorte d'interpolation. Une montée de 1 dans le processus des hauteurs est conservée dans le tour : elle correspond à la visite du premier fils. Un pas de $k \leq 0$ dans le processus des hauteurs correspond à $-k + 1$ pas de -1 dans le tour, suivi d'une montée de 1 pas (deuxième colonne de la Figure 11). Le fait que le processus

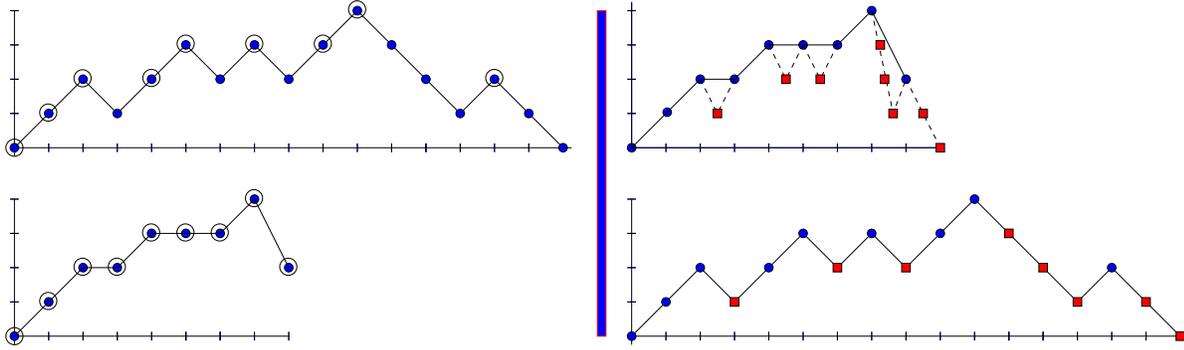


Figure 11: Passage du tour au processus des hauteurs sur la première colonne, et la réciproque dans la deuxième.

des hauteurs et le tour ont la même asymptotique (à un facteur temporel près) est une propriété qui s'étend très nettement hors du cadre des arbres de GW conditionnés par la taille (et ceci est souvent utilisé dans la littérature) : on a

Lemme 1 Soit A_n un ensemble d'arbres enracinés ayant n arêtes, muni d'une loi de probabilité μ_n . On note \tilde{H}_n (resp. \tilde{V}_n) le processus des hauteurs associé (resp. tour associé). Supposons qu'il existe une suite de réels $(c_n)_{n \geq 0}$ tels que $c_n \rightarrow +\infty$, $c_n = o(n)$, et

$$\left(\frac{\tilde{H}_n(nt)}{c_n} \right)_{t \in [0,1]} \xrightarrow[n]{\text{faible}} \left(h(t) \right)_{t \in [0,1]}$$

dans $C[0,1]$ alors

$$\sup_{t \in [0,1]} \frac{|\tilde{V}_n(2nt) - \tilde{H}_n(nt)|}{c_n} \xrightarrow[n]{\text{proba.}} 0.$$

En particulier, on a

$$\left(\frac{\tilde{V}_n(2nt_1)}{c_n}, \frac{\tilde{H}_n(nt_2)}{c_n} \right)_{(t_1, t_2) \in [0,1]^2} \xrightarrow[n]{\text{faible}} \left(h(t_1), h(t_2) \right)_{(t_1, t_2) \in [0,1]^2}.$$

(On trouve une preuve de ce lemme dans [10].) Réciproquement, la convergence du tour sous les mêmes conditions implique également la convergence du processus des hauteurs. Je vais m'attarder sur l'explication du rapport de $\sigma^2/2$ entre le processus des hauteurs et la PPP.

Lien entre la PPP et le processus des hauteurs

On note v_0, v_1, \dots, v_{n-1} les nœuds d'un arbre de taille n , ordonnés selon l'ordre lexicographique, et S_n la PPP correspondante. On dit que $S_n(j)$ est un minimum à droite pour S_n sur $\llbracket 0, l-1 \rrbracket$ si

$$\min_{0 \leq k \leq l-j} \{S_n(j+k)\} = S_n(j).$$

Lemme 2 (Le Gall & Le Jan [80]) *Pour tout $l \in \llbracket 0, n-1 \rrbracket$ et $j \in \llbracket 0, l-1 \rrbracket$, $S_n(j)$ est un minimum à droite pour S_n sur $\llbracket 0, l-1 \rrbracket$ si et seulement si v_j est un ancêtre de v_l .*

Lorsque que $S_n(j)$ est un minimum à droite pour S_n sur $\llbracket 0, l-1 \rrbracket$, on dira que v_j est un minimum à droite sur $\llbracket 0, l-1 \rrbracket$.

L'explication de ce lemme permet également de comprendre la terminologie "PPP". Nous allons voir la marche S_n comme étant la longueur d'une file d'attente dans un système *dernier arrivé - premier servi* (autrement dit, une pile), dans laquelle un client est servi par unité de temps, et dans laquelle les clients de la file sont les nœuds (v_0, \dots, v_{n-1}) . On suppose que

- La racine arrive dans la file au temps 0.
- Les fils de v_i arrivent au temps $i+1$ du plus à droite au plus à gauche de sorte que, le plus à gauche arrivant le dernier sera servi le premier.

Au temps i , le client en tête de queue est servi et s'en va au temps $i+1$.

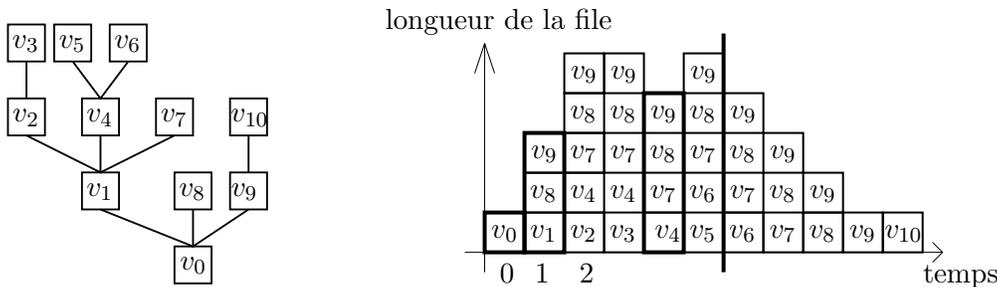


Figure 12: Évolution de la file LIFO associée à un arbre. Le nœud 5 est servi au temps 5 (ce qui est normal, par construction). Dans l'arbre, sa profondeur est 3 puisqu'il a 3 ancêtres v_4 , v_1 et v_0 . Les temps de service de ses ancêtres sont exactement les minima à droite avant le temps 5.

La taille de la queue en fonction du temps est $(S_n(i) + 1)_{i=0, \dots, n}$; en effet, lorsque le client i est servi (-1), il est remplacé par ses fils ($+Y_i$); de plus au temps 0, la taille de la file est 1.

Par construction, v_k est servi au temps k . Durant la visite du sous-arbre $t(v_k)$ enraciné en v_k , la longueur de la pile S_n est supérieure où égale à $S_n(k)$ puisque les nœuds qui étaient là au temps k sont encore là, excepté v_k , et au moins un nœud de $t(v_k)$ est présent jusqu'à la fin de la visite de $t(v_k)$. Une fois le sous-arbre $t(v_k)$ visité, S_n vaut exactement $S_n(k) - 1$, car elle contient les mêmes nœuds qu'au temps k , v_k excepté. Ainsi v_k est un minimum à droite pour ses descendants, et pas plus. Ceci prouve le lemme. \square

La profondeur de v_l dans l'arbre est égale au nombre d'ancêtres de v_l et donc, est égale au nombre de minima à droite de S_n sur $\llbracket 0, l-1 \rrbracket$. Notons

$$A(l) = \{j \mid 0 \leq j \leq l-1, \min\{S_n(j+k), 0 \leq k \leq l-j\} = S_n(j)\}, \quad (5)$$

l'ensemble des indices j pour lesquels v_j est un minimum à droite sur $\llbracket 0, l-1 \rrbracket$. On obtient

$$h(l) = \#A(l). \quad (6)$$

Remarquons encore qu'à l'aide de S_n , nous pouvons localiser les ancêtres de v_k mais aussi donner le nombre de fils de ses ancêtres et également dans quelle lignée se trouve v_k .

SUR LE TYPE DES ANCÊTRES : Notons

$$\begin{aligned} a_k(v_l) &= \#\{j \in A(l) \text{ t.q. } S_n(j+1) - S_n(j) = k-1\} \\ a_{k,i}(v_l) &= \#\{j \in A(l) \text{ t.q. } S_n(j+1) - S_n(j) = k-1, S_n(j^*) - S_n(j) = k-i\}. \end{aligned}$$

où j^* désigne le plus petit élément de $A(l)$ supérieur à j (et lorsqu'il n'y a pas de tel élément, on prend $j^* = l$).

$a_k(u)$ est le nombre d'ancêtres de u ayant k fils. On dira de k -ancêtres de u .

$a_{k,i}(u)$ est le nombre de k -ancêtres de u dont u est un descendant du i -ème. On parle alors de k, i -ancêtres de u .

Par abus de notation, notons $h(u)$ la profondeur du nœud u . On a

Théorème 2 [4] *i)* Soit $\beta > 0$. Il existe $\gamma > 0$ t.q.

$$\mathbb{P}_n \left(\sup_{u,k} |a_k(u) - k p_k h(u)| > \gamma n^{1/4} \ln n \right) = o(n^{-\beta}).$$

ii) Soit $\beta > 0$. Il existe $\gamma > 0$ t.q.

$$\mathbb{P}_n \left(\sup_{u,k,j} |a_{k,j}(u) - p_k h(u)| > \gamma n^{1/4} \ln n \right) = o(n^{-\beta}).$$

Commentaires

- On peut calculer des constantes γ valables explicitement.
- La hauteur des arbres de GW sous \mathbb{P}_n est de l'ordre de $n^{1/2}$ et la proportion des nœuds à profondeur supérieure à $n^{1/2-\varepsilon}$ tend vers 1. Ainsi, ces résultats ne sont pas triviaux ! Ces résultats disent que, grossièrement, le nombre de k -ancêtres de u est proportionnel à $h(u)$ et que $a_{k,j}$ vaut à peu près a_k/k . On a alors, comme conséquences :

$$\frac{\max_u a_k(u)}{\sqrt{n}} \xrightarrow[n]{(d)} \frac{2kp_k}{\sigma} \max e \quad \text{et} \quad \frac{\max_u a_{k,j}(u)}{\sqrt{n}} \xrightarrow[n]{(d)} \frac{2p_k}{\sigma} \max e.$$

- Ceci permet de montrer aussi que “la (ou les) plus grand(es) branche(s)” dans un arbre de GW sous \mathbb{P}_n , contient à des miettes près, le plus grand nombre de nœuds de chaque type. Ceci généralise un résultat de Flajolet & Odlyzko [59] qui affirmait que dans les arbres binaires,

$$n^{-1/2} |\max_u a_{1,2}(u) - \max_u a_{2,2}(u)| \xrightarrow[n]{proba.} 0 \quad \text{et} \quad n^{-1/2} |2 \max_u a_{1,2}(u) - \max_u h(u)| \xrightarrow[n]{proba.} 0.$$

- Ces résultats rappellent la loi de reproduction de la branche infinie dans un arbre de GW conditionné par la non extinction. La raison intuitive est que l'on s'intéresse à des longues branches (de l'ordre de \sqrt{n}), et donc en quelque sorte, à des branches conditionnées à ne pas s'éteindre (voir Bennes & Kersting [30], Lyons & al. [88]).

Éléments de preuve des théorèmes

Tous les résultats annoncés dans le paragraphe précédent reposent sur des propriétés concernant les minima à droite de l'excursion S_n (la PPP), qui est une marche aléatoire conditionnée par (C') .

EFFET DU CONDITIONNEMENT PAR (C') .

Les excursions ont une masse de l'ordre de $n^{-3/2}$ sur l'espace des marches aléatoires (non conditionnées). Il est alors trivial de prouver qu'un événement de probabilité p_n sur l'espace des marches aléatoires aura une probabilité $O(p_n n^{3/2})$ sur l'espace des excursions (sous la loi conditionnelle).

Cette remarque très simple, déjà utilisée dans [14], est rétrospectivement très importante. De nombreuses fonctionnelles des marches aléatoires (le nombre de pas +2 par exemple) présentent des phénomènes de concentration. Ces fonctionnelles seront encore concentrées sur l'ensemble des excursions, sous la loi conditionnelle. Ceci fournit toute une classe de résultats facilement démontrables sur l'espace des excursions.

Ainsi, la démarche que nous avons systématisée est la suivante : nous montrons des résultats de concentration pour une fonctionnelle F des excursions en prouvant un résultat similaire pour la même fonctionnelle F des marches aléatoires, ou plus exactement, pour une fonctionnelle G dont la restriction à l'ensemble des excursions coïncide avec F (par exemple, on peut déduire que la fonctionnelle " $F = \max$ " est concentrée sur l'ensemble des excursions du fait que la fonctionnelle $G = \max - \min$ est concentrée sur l'ensemble des marches).

Pour prouver le Théorème 1, on montre le résultat plus fort suivant :

Théorème 3 [5] *Il existe une constante $c > 0$ telle que pour tout $\nu > 0$,*

$$\mathbb{P} \left(\sup_{0 \leq l \leq n} \left\{ |S_n(l) - h_n(l) \frac{\sigma^2}{2}| \geq n^{1/4+\nu} \right\} \right) = O(\exp(-cn^\nu)).$$

DÉCONDITIONNEMENT : on considère une marche aléatoire $(W(i))_{i \geq 0}$ dont les incréments, les $w(i)$, sont i.i.d. et sont de loi μ avec $(\mu_j)_{j \geq -1} = (p_{j+1})_{j \geq -1}$. Ainsi, S_n à même loi que $(W(i))_{0 \leq i \leq n}$ conditionnée par (C') . On va étudier sur la marche W la variable $RM(l)$ qui donne le nombre de minima à droite de $(W(i))_i$ sur $\llbracket 0, l-1 \rrbracket$:

$$RM(l) = \#\{j | 0 \leq j \leq l-1, \min_{0 \leq k \leq l-j} \{W(j+k)\} = W(j)\}. \quad (7)$$

Le résultat sur les marches dont nous nous servons pour prouver le Théorème 3 est le suivant :

Lemme 3 [5] *Il existe une constante $c > 0$ telle que pour tout $\nu > 0$,*

$$\mathbb{P} \left(\sup_{0 \leq l \leq n} \left\{ |W(l) - \min\{W(j), 0 \leq j \leq l\} - RM(l) \frac{\sigma^2}{2}| \geq n^{1/4+\nu} \right\} \right) = O(\exp(-cn^\nu)).$$

Puisque $\min\{W(j), 0 \leq j \leq l\} = 0$ sur l'espace des excursions, la fonctionnelle

$$G(W) = |W(l) - \min\{W(j), 0 \leq j \leq l\} - RM(l) \frac{\sigma^2}{2}|$$

coïncide avec

$$F(S_n) = |S_n(l) - h_n(l) \frac{\sigma^2}{2}|$$

sur l'espace des excursions. Ainsi,

$$\mathbb{P} \left(\sup_{0 \leq l \leq n} \left\{ |S_n(l) - h_n(l) \frac{\sigma^2}{2}| \geq n^{1/4+\nu} \right\} \right) = O(\exp(-cn^\nu) n^{3/2}),$$

et le Théorème 3 est donc une conséquence de ce lemme.

RETOURNEMENT DE LA MARCHÉ.

Pour prouver le Lemme 3, nous procédons à un retournement de la marche W (voir Fig. 13) afin de pouvoir utiliser la propriété de Markov forte des marches aléatoires. On considère donc $(W^\bullet(j))_{j \in \llbracket 0, l \rrbracket}$ la marche aléatoire définie par :

$$W^\bullet(j) = W(l) - W(l - j) \text{ pour tout } j \in \llbracket 0, l \rrbracket. \quad (8)$$

Les incréments de $(W^\bullet(i))_{i \geq 0}$ ont aussi pour distribution $(\mu_j)_{j \geq -1}$.

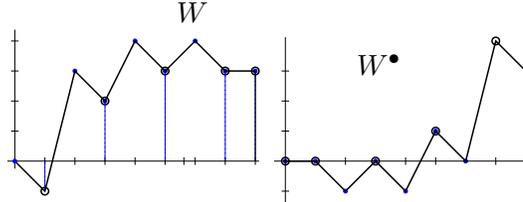


Figure 13: Retournement de la marche. Le graphe de W^\bullet est le symétrique du graphe de W par rapport au centre $(l/2, W(l)/2)$. Les minima à droite sont marqués sur W et deviennent des records sur W^\bullet .

Soit $R^\bullet(l)$ le nombre de records (au sens large) de la trajectoire $(W^\bullet(j))_{j \in \llbracket 0, l \rrbracket}$. Par (8), on obtient facilement

$$\left| \max_{0 \leq j \leq l} \{W^\bullet(j)\} - R^\bullet(l) \frac{\sigma^2}{2} \right| = \left| W(l) - \min_{0 \leq i \leq l} \{W(i)\} - RM(l) \frac{\sigma^2}{2} \right|.$$

L'étude des minima à droite de W se ramène donc à l'étude des records de W^\bullet , ce qui est classique :

LES VARIABLES D'ÉCHELLES

Les temps d'accroissement de W^\bullet , les $0 = \tau_0 < \tau_1 < \tau_2 < \dots$ sont définis par :

$$\tau_k = \inf \{j > \tau_{k-1}, W^\bullet(j) \geq W^\bullet(\tau_{k-1})\}, \text{ pour } k \geq 1. \quad (9)$$

La variable aléatoire $w(\tau_m)$, pour $m \geq 1$ est l'incrément correspondant au m -ème temps d'accroissement ; on appelle ces variables les incréments d'échelles. La loi de $w(\tau_m)$ ne dépend pas de m . Notons aussi que $w(\tau_1)$ n'est pas en général égal à l'accroissement de record $W^\bullet(\tau_1)$.

Lemme 4 Pour tout $k \in \mathbb{N}$, on a

$$\mathbb{P}(W^\bullet(\tau_1) = k) = \mathbb{P}(w(1) \geq k), \quad (10)$$

et donc

$$\mathbb{E}(W^\bullet(\tau_1)) = \sigma^2/2. \quad (11)$$

La loi de $w(\tau_1)$ est donnée par

$$\mathbb{P}(w(\tau_1) = k) = (k+1)\mu_k \text{ pour } k \geq 0. \quad (12)$$

Pour $0 \leq r \leq k$, si $\mu_k > 0$,

$$\mathbb{P}(W^\bullet(\tau_1) = r \mid w(\tau_1) = k) = \frac{1}{k+1}. \quad (13)$$

Ces résultats se montrent par une décomposition idoine de chaque trajectoire favorable ; une preuve partielle se trouve dans Feller 1 (voir [5, 4]).

Il est alors facile de montrer le

Lemme 5 [5] *Pour tout $\nu > 0$, il existe $\gamma > 0$ tel que pour tout n assez grand et tout $l \in \llbracket 0, n \rrbracket$,*

$$\mathbb{P} \left(\left| \max_{0 \leq j \leq l} \{W^\bullet(j)\} - \frac{\sigma^2}{2} R^\bullet(l) \right| \geq n^{1/4+\nu} \right) \leq e^{-\gamma n^\nu}. \quad (14)$$

Ceci traduit le fait que le maximum est proportionnel (à des miettes près) au nombre de records. Le facteur $\sigma^2/2$ est simplement la moyenne de chaque accroissement de record. Le contrôle exponentiel de la probabilité est une conséquence de résultats usuels de concentration (voir Petrov [94] par exemple). Par retournement de la marche, on montre le Lemme 3, puis par conditionnement le Théorème 3. Les autres propriétés des variables d'échelles sont utilisées pour le Théorème 2.

Commentaires

- Le facteur $\frac{\sigma^2}{2}$ est donc la moyenne de $W^\bullet(\tau_1)$.
- Prenons un jeu de moyenne 0, avec mise initiale 1 et gain aléatoire. On commence avec une fortune nulle ; puis on s'intéresse au temps de retour dans \mathbb{N} après le temps 0. La probabilité que la fortune devienne positive lors d'un gain de k est $k + 1$ fois la probabilité de gagner k en un coup. La formule (13) précise alors qu'à ce moment, la fortune est uniformément distribuée sur $\llbracket 0, k \rrbracket$.
- D'après (10), si $w(1)$ possède des moments exponentiels $W^\bullet(\tau_1)$ aussi. Ceci est important pour les bornes exponentielles des probabilités comme dans la formule (14).
- L'existence de moments exponentiels pour $w(1)$ (ou de manière équivalente pour la loi de reproduction) n'est pas nécessaire pour montrer que $n^{-1/2} \|\sigma^2 h_n / 2 - S_n\|_\infty \xrightarrow[n]{proba.} 0$. Une analyse grossière montre que des moments d'ordre 9 suffisent. Des simulations suggèrent que sans l'existence de moments d'ordre 3, ce n'est peut être plus le cas.
- Avec M. Drmota [7], nous avons montré que la convergence de $V_n(2n.) / \sqrt{n}$ vers l'excursion brownienne était plus forte que la convergence faible (les fonctionnelles sous-polynomiales en $\|V_n / \sqrt{n}\|_\infty$ sont équi-intégrables et convergent en loi). Ce résultat est générique en combinatoire où lors de la preuve de la convergence faible, la tension d'une suite de processus est montrée à l'aide du contrôle des moments de $|x_n(s) - x_n(t)|$. Dans le cas des arbres, ceci permet de montrer que la convergence faible de la hauteur des arbres vers le maximum de l'excursion brownienne est également une convergence des moments (ceci étant connu et dû à Flajolet & Odlyzko [59]) ; on obtient également la convergence des moments du déplacement total.

1.5 Arbre abstrait

Dans cette section est donnée *in extenso* une définition d'arbre abstrait légèrement différente de celles que l'on trouve dans la littérature ; je renvoie ici à Aldous & al. [23] pour un historique de l'objet "arbre continu" (on pourra voir aussi directement les œuvres d'Aldous, Duquesne, Le Gall, Le Jan, Evans, Winter, en particulier [18, 19, 20, 50, 55, 79, 51, 81, 97]).

Dans la Section 3, avec A. Mokkadem, nous montrons qu'une quadrangulation s'obtient en "collant les nœuds d'un premier arbre sur les coins d'un deuxième arbre". Afin d'obtenir une notion de carte abstraite comprenant aussi bien les cartes discrètes que continues, nous avons absolument besoin qu'un arbre détermine l'ensemble de ses coins. Le tour permet de distinguer seulement les nœuds de degré différent de 1 : ils correspondent à des points étant des maxima locaux ou des minima locaux dans le tour. Ainsi, on ne peut pas en général reconnaître un nœud de l'arbre s'il n'est pas *donné extrinsèquement*. Ceci amène Pitman [97] à introduire une notion d'arbre réduit, où les nœuds de degré 1 ne sont pas considérés comme des nœuds mais comme des points d'arrêt. Nous avons dit plus haut que l'on pouvait reconstruire l'arbre à partir de son tour, mais c'était à n fixé avec un arbre (discret donc) dont les longueurs d'arêtes étaient égales à 1 ou à $1/\sqrt{n}$ (dans la version normalisée). Dans le cas général, le tour de l'arbre ne suffit pas à retrouver ses nœuds.

Pour pallier cet inconvénient nous allons définir un arbre comme un couple formé d'une fonction (l'analogie du tour) et d'une mesure "marquant" les nœuds ou plus exactement les coins. La définition des arbres que l'on donne permet de montrer la convergence des arbres de GW normalisés vers un arbre continu (codé par l'excursion brownienne et la mesure de Lebesgue sur $[0,1]$) à l'aide de la convergence du tour vers l'excursion brownienne. La distance sur l'espace des arbres que l'on construit tient compte de la distance entre les tours et de la répartition "de la masse des coins" dans les arbres.

Espace des arbres

Soit $C^+[0,1]$ l'ensemble des fonctions g continues de $[0,1]$ dans \mathbb{R}^+ telles que $g(0) = g(1) = 0$. À toute fonction g de $C^+[0,1]$ on va associer une structure arborescente codée par g de la même manière qu'un tour d'arbre discret code un arbre discret. On définit la relation d'équivalence \sim_g dans $[0,1]$ par

$$x \sim_g y \iff g(x) = g(y) = \check{g}(x,y),$$

où $\check{g}(x,y) = \inf\{g(u), u \in [x \wedge y, x \vee y]\}$.

Notons E_g l'ensemble des classes d'équivalence modulo \sim_g . L'ensemble E_g sera un peu plus loin l'ensemble des points de l'arbre (un point étant un nœud, ou un "point d'arête"). On note F_g la surjection canonique de $[0,1]$ dans E_g ; c'est-à-dire, pour $x \in [0,1]$, on a $F_g(x) \stackrel{not}{=} \dot{x} = \{y, y \in [0,1], x \sim_g y\}$. La fonction F_g est l'analogie de la FPP définie page 13.

Afin de distinguer intrinsèquement les nœuds, on considère une mesure μ sur $[0,1]$ (ceci induit une mesure sur E_g par F_g); on définit l'ensemble des nœuds grâce à cette mesure. Soit \mathcal{M} l'ensemble des mesures finies sur $[0,1]$ et pour $\mu \in \mathcal{M}$ posons $E_\mu = F_g(\text{support}(\mu))$.

On dit qu'un couple $(g, \mu) \in C^+[0,1] \times \mathcal{M}$ est le codage d'un arbre s'il satisfait la condition suivante :

$$E_g^{(T)} \stackrel{def}{=} \{u \in E_g, \#u \neq 2\} \cup \{\dot{0}\} \subset E_\mu.$$

Cette condition impose de considérer comme nœud (les nœuds seront les éléments de E_μ) au moins tous les points contenant une information topologique : la racine, les feuilles, et les points qui ont un degré différent de 1.

Arbre : définition

L'arbre \mathcal{T} codé par (g, μ) et noté $\mathcal{T} = \text{CTREE}(g, \mu)$, est l'espace métrique totalement ordonné \mathcal{T} défini comme suit :

- L'ensemble des points de \mathcal{T} est $E_{\mathcal{T}} = E_g$. Le parcours en profondeur de \mathcal{T} est la fonction $F_{\mathcal{T}} = F_g$.
- L'ensemble des nœuds est E_μ . La classe $F_{\mathcal{T}}(0) = \dot{0}$ est appelée racine de \mathcal{T} .
- L'ensemble des *coins* autour d'un point \dot{x} est $F_{\mathcal{T}}^{-1}(\dot{x}) \cap [0,1[$ (les coins de \dot{x} sont les temps de visite de \dot{x} par $F_{\mathcal{T}}$; ils fournissent les renseignements nécessaires pour reconstituer la structure arborescente de l'arbre). Le coin 0 est le coin racine.
- Distance sur \mathcal{T} : Soit x et y des représentants de \dot{x} et \dot{y} . La classe (modulo g) de z définie par $z \in [x,y]$ et $g(z) = \check{g}(x,y)$ ne dépend pas des représentants de \dot{x} et \dot{y} choisis. Le point z est appelé plus *profond ancêtre commun* de \dot{x} et \dot{y} . La distance entre \dot{x} et \dot{y} est définie par $d_{\mathcal{T}}(\dot{x}, \dot{y}) = d_{\mathcal{T}}(x, y) = g(x) + g(y) - 2g(z) = g(x) + g(y) - 2\check{g}(x, y)$.

Il reste à voir que \mathcal{T} présente toutes les caractéristiques d'un arbre.

Structure arborescente de \mathcal{T} et ordre sur \mathcal{T}

- 1) Par la définition de $d_{\mathcal{T}}$, la fonction $F_{\mathcal{T}}$ est continue (de $[0, 1]$ dans $E_{\mathcal{T}}$).
- 2) Ainsi $E_{\mathcal{T}}$ est un espace métrique compact, connexe par arcs : il est paramétré par $[0, 1]$.
- 3) Le tour de \mathcal{T} est g (c'est-à-dire $g(x) = d_{\mathcal{T}}(F_{\mathcal{T}}(x), F_{\mathcal{T}}(0))$).
- 4) $\#F_{\mathcal{T}}^{-1}(u) \cap [0, 1[$ est appelé le degré total de u (que l'on note $\deg(u)$).
- 5) Soit $\dot{x} \in E_{\mathcal{T}}$, on pose $\underline{\rho}(\dot{x}) = \inf\{y \in [0, 1], F_{\mathcal{T}}(y) = \dot{x}\}$ et $\bar{\rho}(\dot{x}) = \sup\{y \in [0, 1], F_{\mathcal{T}}(y) = \dot{x}\}$. L'intervalle $[\underline{\rho}(\dot{x}), \bar{\rho}(\dot{x})]$ est appelé sous-arbre enraciné en \dot{x} .
- 6) Si $(\dot{x} = \dot{y}, \dot{z} \neq \dot{x}, \text{ et } x < z < y)$ alors $\dot{z} \subset [x, y]$. Dans ce cas, on dit que \dot{z} est un descendant de \dot{x} (ou que \dot{x} est un ancêtre de \dot{z}).
- 7) Si $[\underline{\rho}(\dot{x}), \bar{\rho}(\dot{x})] \cap [\underline{\rho}(\dot{y}), \bar{\rho}(\dot{y})] \neq \emptyset$ alors $[\underline{\rho}(\dot{x}), \bar{\rho}(\dot{x})] \subset [\underline{\rho}(\dot{y}), \bar{\rho}(\dot{y})]$ ou $[\underline{\rho}(\dot{y}), \bar{\rho}(\dot{y})] \subset [\underline{\rho}(\dot{x}), \bar{\rho}(\dot{x})]$. Dans le premier cas, on dit que \dot{x} est un descendant de \dot{y} , dans le second que \dot{y} est un descendant de \dot{x} .
- 8) La relation d'ordre est définie ainsi :

$$\dot{x} \preccurlyeq \dot{y} \Leftrightarrow \underline{\rho}(\dot{x}) \leq \underline{\rho}(\dot{y}).$$

On dit que le nœud \dot{x} est à gauche du nœud \dot{y} (en fait, ceci signifie que par le parcours en profondeur on visite un coin de \dot{x} avant de visiter un coin de \dot{y}). Ceci fait de $E_{\mathcal{T}}$ un ensemble totalement ordonné. En particulier, si \dot{x} est un ancêtre de \dot{y} alors $\dot{x} \preccurlyeq \dot{y}$.

9) Si $x_1 < x_2 < x_3$ sont des représentants de \dot{x} , $[x_1, x_2]$ et $[x_2, x_3]$ sont des sous-arbres de $[x_1, x_3]$ et on dit que $[x_1, x_2]$ est à gauche de $[x_2, x_3]$.

10) Branches et géodésiques : l'unique géodésique \mathcal{S} entre un point u et la racine est la courbe continue paramétrée comme suit. Soit x tel que $\dot{x} = u$ et pour $s \in [0, g(x)]$ posons $m(s) = \sup\{y \in [0, x], g(y) = s = \inf_{y \leq z \leq x} g(z)\}$; la fonction $s \mapsto F_{\mathcal{T}}(m(s))$ est une bijection continue entre $[0, g(x)]$ et \mathcal{S} . On peut voir qu'entre tout couple de points (\dot{x}, \dot{y}) , il n'y a qu'une seule géodésique paramétrée par $[\dot{g}(x, y), g(x)] \cup [\dot{g}(x, y), g(y)]$. Une géodésique entre un point et la racine est appelée branche.

11) \mathcal{T} n'a pas de cycle (ceci se montre aisément).

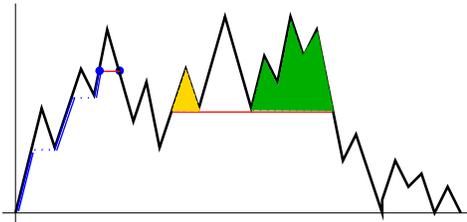


Figure 14: Une fonction f considérée comme tour d'arbre. Les points x, y, z et t codent quatre coins du même nœud. Les intervalles $[x, y], [y, z]$ et $[z, t]$ codent 3 sous-arbres enracinés en \dot{x} .

Ensemble des arbres

Soit Γ l'ensemble des arbres. Une métrique sur Γ est donnée par d_{Γ} définie comme suit : soient $\mathcal{T}_1 = \text{CTREE}(g_1, \mu_1)$ et $\mathcal{T}_2 = \text{CTREE}(g_2, \mu_2)$ deux éléments de Γ

$$d_{\Gamma}(\mathcal{T}_1, \mathcal{T}_2) = \|g_1 - g_2\|_{\infty} + d_{\mathcal{M}}(\mu_1, \mu_2)$$

où

$$\|g_1 - g_2\|_{\infty} = \sup\{|g_1(x) - g_2(x)|, x \in [0, 1]\}$$

et

$$d_{\mathcal{M}}(\mu_1, \mu_2) = \sup_{x \in \mathbb{R}} |C_{\mu_1}(x) - C_{\mu_2}(x)|$$

où C_μ est la fonction de répartition de μ (c'est-à-dire $C_\mu(x) = \mu] - \infty, x]$). Avec cette distance on peut définir une notion de convergence et de convergence faible dans Γ .

Arbres discrets vus comme arbres abstraits

Dans le cas discret, afin d'obtenir la convergence du tour, on normalise le processus V_n :

$$v_n(t) = \frac{V_n(2(n-1)t)}{\sqrt{n}} \quad \text{pour } t \in [0, 1].$$

Ceci revient à considérer un arbre dont les longueurs d'arêtes seraient $n^{-1/2}$ et que l'on visiterait à la vitesse de $2(n-1)$ arêtes par unité de temps. On considère maintenant une mesure sur $[0, 1]$ dont l'image du support par F_g est l'ensemble des nœuds de l'arbre. Il est légitime de considérer la mesure suivante :

$$\mu_n = \frac{1}{2(n-1)} \sum_{k=0}^{2n-3} \delta_{k/(2(n-2))}.$$

On a mis une masse de $(2(n-1))^{-1}$ sur chacun des temps de visite des nœuds, excepté sur la racine qui est visitée une fois de plus que son degré. On a $E_{\mu_n} = \{\hat{x}, x = k/(2(n-1)), k \in \llbracket 0, 2(n-1) \rrbracket\}$. On aurait aussi pu donner une masse de $1/n$ à chaque nœud sans que cela ne change l'asymptotique (on peut répartir la masse $1/n$ affectée à un nœud de manière arbitraire).

Les arbres de GW conditionnés par $\{X = n\}$, normalisés, sont du type $\text{CTREE}(v_n, \mu_n)$ et appartiennent à Γ .

Arbres continus et convergence

On a vu que $v_n \xrightarrow[n]{faible} \frac{2}{\sigma}e$. Il est facile de voir que $\mu_n \xrightarrow[n]{faible} \text{Leb}_{|[0,1]}$ la mesure de Lebesgue sur $[0, 1]$; ainsi

$$C_{\mu_n} \xrightarrow[n]{unif.} C_{\text{Leb}_{|[0,1]}},$$

la fonction de répartition de la loi uniforme $\mathcal{U}[0, 1]$. Dans ce cas, la convergence est déterministe.

On appelle *arbre continu aléatoire*, l'arbre $\text{CTREE}(\frac{2}{\sigma}e, \text{Leb}_{|[0,1]})$. Ainsi, puisque le support de la mesure est $[0, 1]$ on considère que tous les points sont coins et donc que tous représentent des nœuds.

Il est maintenant facile de montrer la convergence des arbres de GW conditionnés par la taille vers les arbres continus aléatoires :

Théorème 4 *La convergence faible suivante a lieu dans Γ :*

$$\text{CTREE}(v_n, C_{\mu_n}) \xrightarrow[n]{faible} \text{CTREE}\left(\frac{2}{\sigma}e, \text{Leb}_{|[0,1]}\right).$$

2 Des marches branchantes

2.1 Premiers travaux

Nous nous intéressons ici aux marches branchantes conditionnées par la taille de l'arbre sous-jacent. Ce modèle a été peu étudié dans la littérature ; on trouvera des résultats sur le maximum de ces marches dans Kesten [74] et Durrett & al. [53].

Avec A. Mokkadem [3], nous avons considéré le modèle suivant :

- Un arbre T est choisi uniformément dans l'ensemble Ω_n des arbres planaires ayant n arêtes (voir p.17).
- Sur chaque branche de l'arbre T , nous construisons une marche aléatoire. À chaque nœud u est associée une variable aléatoire réelle $y(u)$, que l'on appelle *marque* de u . On pose $y(\emptyset) = 0$ et on suppose que les variables aléatoires $y(u)$ pour $u \neq \emptyset$ sont i.i.d. et indépendantes de l'arbre T et que

$$\mathbb{E}(y(u)) = 0, \quad \text{Var}(y(u)) = 1, \quad \text{il existe } p > 6 \text{ t.q. } \mathbb{E}(|y(u)|^p) < +\infty. \quad (15)$$

Soit u un nœud, $h(u)$ sa profondeur et $(u_0 = \text{racine}, u_1, \dots, u_{h(u)} = u)$ la géodésique entre la racine et u . On associe à u la trajectoire finie $\Phi_u = (\Phi_u(j))_{j \in \llbracket 0, h(u) \rrbracket}$ définie par

$$\Phi_u(0) = 0, \quad \Phi_u(j) = \sum_{i=1}^j y(u_i), \text{ pour } j \in \llbracket 1, h(u) \rrbracket.$$

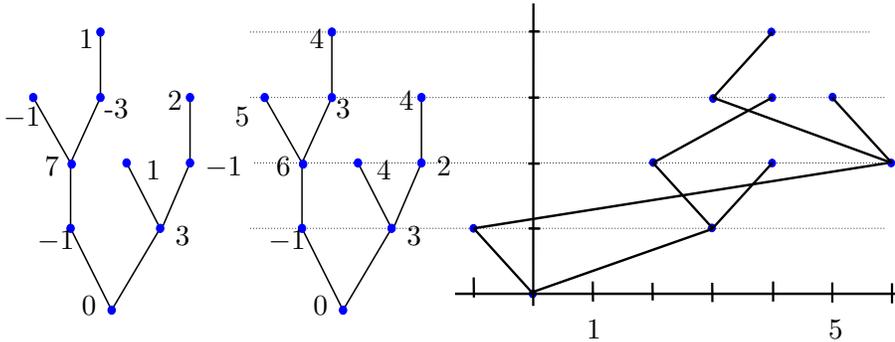


Figure 15: Un arbre marqué, son arbre étiqueté associé et la marche branchante associée

Par analogie avec le serpent brownien, le serpent discret sera défini un peu plus loin comme étant moralement la liste des marches Φ_u ordonnées selon la FPP : $(\Phi_{F(k)})_{k=0, \dots, 2n}$ (on rappelle que $F(k)$ est le nœud visité par la mouche au temps k).

On appelle *étiquette* du nœud u , la valeur terminale de la marche Φ_u , à savoir $\Phi_u(h(u))$ (voir Fig. 15). Constatons que la donnée de T et de la liste des étiquettes permet de reconstituer entièrement l'arbre marqué et également la liste des marches Φ_u . Ceci jouera un rôle important par la suite.

Notre but ici est de montrer que sous certaines conditions concernant les variables $y(u)$, le serpent discret converge vers le serpent brownien de processus de durée de vie l'excursion brownienne, dont la définition est rappelée un peu plus loin. Pour cela, nous allons utiliser des codages naturels (voir Fig. 16) :

- l'arbre sous-jacent T est codé par son tour V_n et par la FPP, la fonction F .
- On code les valeurs terminales des marches $\Phi_u(h(u))$ grâce au processus R_n défini par

$$R_n(k) = \Phi_{F(k)}(h(F(k))).$$

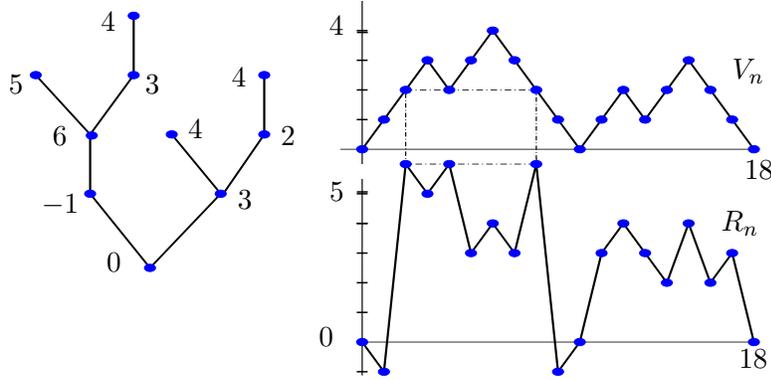


Figure 16: Codage d'un arbre étiqueté par deux processus compatibles : le tour du serpent. Le petit rectangle en pointillés illustre le fait que R_n est constant sur les nœuds de T

Revenons à l'histoire de la mouche marchant autour de l'arbre. Le processus V_n donne l'ordonnée de la mouche en fonction du temps. Si l'on considère les marques $y(u)$ comme des déplacements horizontaux, le processus $R_n(k)$ donne l'abscisse de la mouche en fonction du temps.

On appelle le processus (R_n, V_n) le *tour du serpent discret* et le processus R_n *tête du serpent*, ou *processus des étiquettes*. Pour montrer un comportement asymptotique pour les serpents discrets, nous les plongeons dans un espace de fonctions, dans lequel la convergence va avoir lieu.

2.1.1 Serpents discrets et continus

On interpole linéairement V_n et R_n entre les points d'abscisses entières (on appelle les processus obtenus par les mêmes noms). Par analogie avec le serpent brownien, on appelle serpent discret le processus

$$(W_n(x, t), V_n(x))_{x \in [0, 2n], t \in [0, \infty)}$$

où pour tout $x \in [0, 2n]$, $W_n(x, \cdot)$ est un processus continu, associé au point de T visité au temps x par la FPP et arrêté au temps $V_n(x)$. Il est défini par :

– Pour $k \in \llbracket 0, 2n \rrbracket$ et $t \in [0, V_n(k)]$, $W_n(k, \cdot)$ est le processus qui interpole linéairement la marche aléatoire $\Phi_{F(k)}$:

$$W_n(k, t) = \Phi_{F(k)}(\lfloor t \rfloor) + \{t\} (\Phi_{F(k)}(\lceil t \rceil) - \Phi_{F(k)}(\lfloor t \rfloor)).$$

– Pour $x \in [0, 2n] \setminus \mathbb{Z}$ et $t \in [0, V_n(x)]$, la trajectoire $W_n(x, \cdot)$ coïncide avec la trajectoire $W_n(\lceil x \rceil, \cdot)$ sur $[0, V_n(x)]$ si $V_n(\lceil x \rceil) \geq V_n(x)$ ou avec $W_n(\lfloor x \rfloor, \cdot)$ sinon :

$$W_n(x, t) = \begin{cases} W_n(\lfloor x \rfloor, t) & \text{si } V_n(\lceil x \rceil) < V_n(\lfloor x \rfloor) \\ W_n(\lceil x \rceil, t) & \text{si } V_n(\lceil x \rceil) > V_n(\lfloor x \rfloor). \end{cases}$$

(Voir Fig. 17)

– Pour $x \in [0, 2n]$ et $t \in]V_n(x), +\infty[$, on pose

$$W_n(x, t) = W_n(x, V_n(x)). \quad (16)$$

On a alors

$$W_n(\lfloor x \rfloor, t) = W_n(\lceil x \rceil, t) \text{ pour } 0 \leq t \leq V_n(\lfloor x \rfloor) \wedge V_n(\lceil x \rceil).$$

Ainsi, le serpent discret apparaît comme étant une famille de processus aléatoires dépendants,

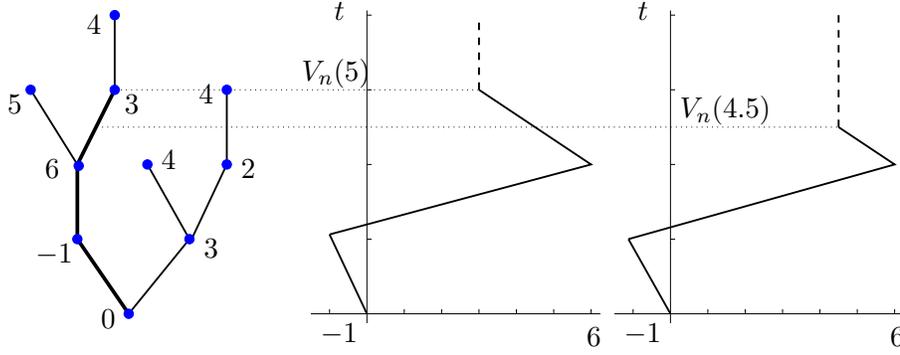


Figure 17: Un arbre étiqueté, puis le graphe de $(W(5, t), t)$ et de $(W(4.5, t), t)$. Il faut remarquer que $(W(5, t), t) = (W(7, t), t)$ et que $(W(4.5, t), t) = (W(7.5, t), t)$ pour tout t

indexée par l'intervalle $[0, 2n]$. Le tour du serpent discret (R_n, V_n) est par nature plus simple même si, évidemment, la donnée du tour permet de reconstituer entièrement le serpent et réciproquement.

NORMALISATIONS : afin d'obtenir des théorèmes limites pour le tour du serpent discret et pour le serpent discret lui-même, on pose

$$v_n(s) = \frac{V_n(2ns)}{\sqrt{n}} \quad \text{pour } s \in [0, 1], \quad r_n(s) = \frac{R_n(2ns)}{n^{1/4}} \quad \text{pour } s \in [0, 1]$$

et

$$w_n(s, t) = \frac{W_n(2ns, t\sqrt{n})}{n^{1/4}} \quad \text{pour } (s, t) \in [0, 1] \times [0, +\infty[.$$

On appelle les processus (w_n, v_n) et (r_n, v_n) le serpent discret normalisé et le tour normalisé. Notons que w_n satisfait la propriété :

$$w_n(s, t) = w_n(s, v_n(s)) \quad \text{pour } t \geq v_n(s).$$

Espace d'état des serpents

On appelle chemin arrêté un couple $(\nu, \zeta) \in C([0, +\infty[, \mathbb{R}) \times [0, +\infty[$ tel que pour tout $t \geq \zeta$ on a $\nu(t) = \nu(\zeta)$. On note \mathcal{W} l'espace des chemins arrêtés. On appelle ζ la durée de vie de (ν, ζ) et $\nu(\zeta)$ le "point terminal", ou étiquette du chemin ν .

On note \mathcal{S} le sous-espace de $C([0, 1], C([0, +\infty[, \mathbb{R})) \times C([0, 1], [0, +\infty[)$ des fonctions¹ (ν, ζ) , "compatibles", c'est-à-dire qui satisfont les trois conditions suivantes :

$$\begin{cases} (i) & (\nu(s, \cdot), \zeta(s)) \in \mathcal{W} \text{ pour tout } s \in [0, 1] \\ (ii) & \zeta(0) = \zeta(1) = 0 \\ (iii) & \text{Pour tout } 0 \leq s \leq s' \leq 1, \nu(s, t) = \nu(s', t) \text{ pour tout } t \leq \check{\zeta}(s, s'). \end{cases} \quad (17)$$

Il faut comprendre les choses comme suit : la fonction ζ appartient à $C^+[0, 1]$, et est donc le tour d'un arbre. Au nœud ayant pour représentant (ou coin) s , on associe le chemin arrêté $\nu(s, \cdot)$; ce chemin est arrêté en $t = \zeta(s)$, profondeur du coin s dans l'arbre. La condition (iii), dite condition de compatibilité, implique que si s et s' représentent le même nœud, $\nu(s, \cdot) = \nu(s', \cdot)$. De plus par (iii), on voit que pour $t < \zeta(s)$, $\nu(s, t)$ représente l'étiquette de l'unique ancêtre de s ayant pour profondeur t . Ainsi, si s et s' ont même ancêtre à profondeur t , $\nu(s, t) = \nu(s', t)$. La raison pour

¹On note $C(A, B)$ l'espace des fonctions continues définies sur A à valeurs dans B .

laquelle le chemin $\nu(s, \cdot)$ est prolongé et non pas tué en $\zeta(s)$ est technique : ainsi, les fonctions $\nu(\cdot, \cdot)$ sont définies sur $[0, 1] \times \mathbb{R}^+$ et il est alors facile de définir une métrique sur \mathcal{S} . On munit l'espace \mathcal{S} de la distance :

$$d_{\mathcal{S}}((\nu_1, \zeta_1), (\nu_2, \zeta_2)) = \max \{ \|\nu_1 - \nu_2\|_{\infty}, \|\zeta_1 - \zeta_2\|_{\infty} \}$$

où

$$\|\nu_1 - \nu_2\|_{\infty} = \sup_{(s,t) \in [0,1] \times [0,+\infty[} |\nu_1(s,t) - \nu_2(s,t)|.$$

Le serpent normalisé $(w_n(s, \cdot), v_n(s'))_{(s,s') \in [0,1]^2}$ est un élément de \mathcal{S} .

Espace d'état du tour

On note \mathbb{T} le sous-espace de $C([0, 1], \mathbb{R}) \times C([0, 1], [0, +\infty[)$ des fonctions (f, ζ) qui satisfont

$$\begin{cases} \zeta(0) = \zeta(1) = 0 \\ \text{pour tout } 0 \leq s \leq s' \leq 1, \quad f(s) = f(s') \quad \text{si} \quad \zeta(s) = \zeta(s') = \check{\zeta}(s, s'). \end{cases} \quad (18)$$

De nouveau, $\zeta \in C^+[0, 1]$ et code donc un arbre. Si s et s' vérifient $\check{\zeta}(s, s') = \zeta(s) = \zeta(s')$, alors s et s' sont deux représentants du même nœud de l'arbre codé par ζ . La condition (18) nous assure que f est bien constant sur les nœuds. Nous prenons comme distance sur \mathbb{T} :

$$d_{\mathbb{T}}((f_1, \zeta_1), (f_2, \zeta_2)) = \max \{ \|f_1 - f_2\|_{\infty}, \|\zeta_1 - \zeta_2\|_{\infty} \}.$$

Par construction, le tour normalisé du serpent (r_n, v_n) est un élément de \mathbb{T} .

Définition du serpent brownien

On renvoie à [81, 51, 105] pour de plus amples renseignements sur le serpent brownien et autres super-processus.

Le serpent brownien de durée de vie ζ (où $\zeta \in C^+[0, 1]$) est le processus $(w(\cdot, \cdot), \zeta(\cdot))$ à valeurs dans \mathcal{S} dont la loi est définie par :

- Pour tout s , $w(s, 0) = 0$.
- La loi conditionnelle de $w(s, \cdot)$ sachant ζ est la loi d'un processus de Markov homogène dont le noyau de transition est décrit par :

Pour $0 \leq s \leq s' \leq 1$,

* $w(s', t) = w(s, t)$ pour tout $t \leq \check{\zeta}(s, s')$

* $(w(s', \check{\zeta}(s, s') + t)_{t \in [0, \zeta(s') - \check{\zeta}(s, s')]} - w(s', \check{\zeta}(s, s'))$ est indépendant de $w(s, \cdot)$ et distribué comme un mouvement brownien réel partant de $w(s, \check{\zeta}(s, s'))$ et arrêté au temps $\zeta(s') - \check{\zeta}(s, s')$ (voir Fig. 18).

Le tour du serpent brownien

Le tour du serpent brownien de durée de vie $v = \sqrt{2}e$ est le processus (r, v) à valeurs dans \mathbb{T} défini par

$$r(s) = w(s, v(s)) \text{ pour } s \in [0, 1].$$

Le processus r sachant v , est un processus gaussien centré de fonction de covariance,

$$\text{cov}(r(s), r(t)) = \check{v}(s, t).$$

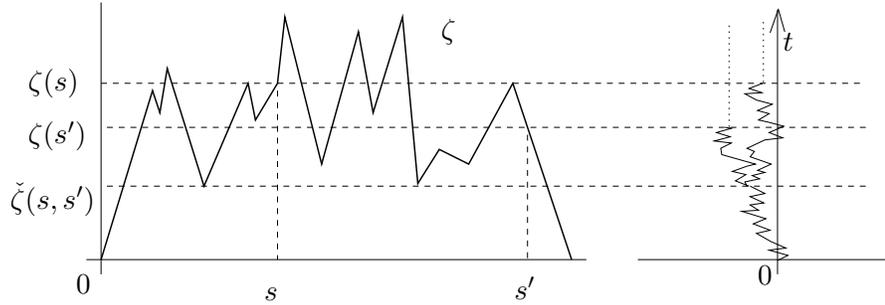


Figure 18: Processus des durées de vie ζ . À droite, représentation schématique de $(w(s, t), t)$ et de $(w(s', t), t)$. Le processus $(t \mapsto w(s, t))$ (resp. $(t \mapsto w(s', t))$) est un mouvement brownien tant que t est inférieur au temps $\zeta(s)$ (resp. $\zeta(s')$). Et $(w(s, t) = w(s', t))$ tant que $t \leq \check{\zeta}(s, s')$.

CONVERGENCE DU TOUR DU SERPENT DISCRET

Théorème 5 [3] *Sous les hypothèses (15), la convergence suivante a lieu dans \mathbb{T}*

$$(r_n, v_n) \xrightarrow[n]{(d)} (r, v),$$

où (r, v) est le tour du serpent brownien de durée de vie $v = \sqrt{2}e$.

La convergence faible de v_n vers v est déjà connue (Section I.1). La convergence des finies dimensionnelles de (r_n, v_n) vers celles de (r, v) est relativement simple à prouver, et est, en gros, une conséquence du TLC. La preuve de la tension de la suite $(r_n)_n$ est plus problématique, d'ailleurs la restriction à l'existence des moments d'ordre $6 + \varepsilon$ provient de là. Nous devons signaler que simultanément, Chassaing & Schaeffer [41] ont obtenu $(r_n, v_n) \xrightarrow[n]{(d)} (r, v)$ pour la topologie de Skohorod, lorsque les marques sont des variables aléatoires i.i.d. uniformes dans $\{+1, -1, 0\}$ (avec une normalisation très légèrement différente pour tenir compte de la variance des marques). Ils montrent également la convergence des moments de $\max_n r_n$.

La convergence du tour du serpent implique la convergence des serpents eux-mêmes. Ceci est expliqué dans la section suivante :

2.1.2 Un homéomorphisme

■ Soit (ν, ζ) un élément de \mathcal{S} (un serpent). On définit $\mathcal{H}(\nu, \zeta) = (f, \zeta)$ par

$$f(s) = \nu(s, \zeta(s)) \text{ pour tout } s \in [0, 1].$$

■ Soit (f, ζ) un élément de \mathbb{T} . On définit $\overline{\mathcal{H}}(f, \zeta) = (\nu, \zeta)$ par

$$\nu(s, t) = f(\rho(s, t, \zeta)) \text{ pour tout } s \in [0, 1] \text{ et } t \geq 0$$

et (voir Fig. 19)

$$\rho(s, t, \zeta) = \begin{cases} \sup\{\alpha \leq s, \zeta(\alpha) = t\} & \text{si } 0 \leq t \leq \zeta(s), \\ s & \text{si } t \geq \zeta(s). \end{cases}$$

Théorème 6 [3] (i) \mathcal{H} est une application bien définie de \mathcal{S} dans \mathbb{T} et $\overline{\mathcal{H}}$ est une application bien définie de \mathbb{T} dans \mathcal{S} .

(ii) \mathcal{H} est bijective et $\mathcal{H}^{-1} = \overline{\mathcal{H}}$.

(iii) \mathcal{H} est un homéomorphisme de \mathcal{S} dans \mathbb{T} .

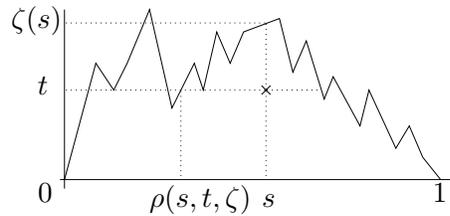


Figure 19: La projection ρ . Le réel $\rho(s, t, \zeta)$ est le représentant le plus à gauche de l'ancêtre de \dot{s} se trouvant à la profondeur t dans l'arbre de tour ζ .

L'application \mathcal{H} associe à un serpent son tour, tandis que l'application $\overline{\mathcal{H}}$ associe à un tour son serpent (celui qui a ce tour). Le théorème dit que ces applications sont bien définies, qu'elles sont bijectives et homéomorphes.

Commentaires

Comme dit plus haut, le tour du serpent et le serpent se caractérisent l'un l'autre. Les fonctions \mathcal{H} et \mathcal{H}^{-1} le montrent et c'est plus ou moins évident. Le fait que les transformations \mathcal{H} et surtout \mathcal{H}^{-1} sont continues est beaucoup moins clair. L'un des intérêts de ce théorème est le suivant : il permet de déduire la convergence d'une suite de serpents de la convergence des tours correspondants ; au niveau des marginales, ceci dit que si la suite de tours $(f_n, \zeta_n)_n$ converge faiblement, alors à s fixé, la suite de processus $\nu_n(s, \cdot)$ converge faiblement. Remarquer que c'est (iii) que l'on utilise pour cela, pas (ii). Le Théorème 6 permet également d'obtenir, par contraction, un principe de grandes déviations pour le serpent brownien (voir [3]) à l'aide de grandes déviations de son tour, résultat obtenu par Serlet [105].

Un corollaire des Théorèmes 5 et 6 est le suivant

Corollaire 1 [3] (*Convergence du serpent discret*) La convergence suivante a lieu dans \mathcal{S}

$$(w_n, v_n) \xrightarrow[n]{(d)} (w, v),$$

où (w, v) est le serpent brownien.

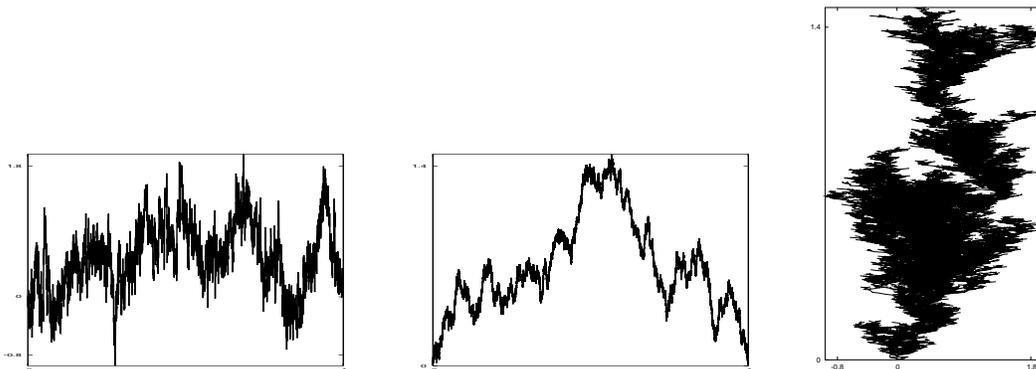


Figure 20: Simulation du tour (r, v) du serpent brownien à l'aide d'une marche branchante de taille 10^5 et représentation du serpent : en ordonnée, profondeur dans l'arbre sous-jacent, en abscisse, processus des étiquettes. Il s'agit plutôt d'une projection du serpent car on a perdu l'ordre des browniens (on a dessiné $\cup_{s \in [0,1]} \{(w(s, t), t), t \leq \zeta(s)\}$).

2.2 Marche branchante : cas général

Gittenberger [63] montra que l'on pouvait étendre les résultats concernant la convergence du tour du serpent – et du serpent lui-même donc, grâce au théorème d'homéomorphisme – à un modèle plus général. Dans son modèle, l'arbre sous-jacent est un arbre de GW critique conditionné par $\{X = n\}$ (la loi de reproduction doit toutefois admettre des moments exponentiels) et les marques sont centrées et possèdent des moments d'ordre $8 + \varepsilon$ (un moment d'ordre $6 + \varepsilon$ était nécessaire dans [3]). Pour cela, il a montré que le lemme de [3] concernant la tension de la famille de processus $(r_n)_n$ s'étendait à son modèle.

Avec S. Janson, nous avons entrepris de comprendre où intervenaient les hypothèses de moments dans la convergence vers le serpent. Dans cette section, j'expose rapidement les résultats obtenus dans [9]. Notre modèle est celui des marches branchantes dont l'arbre sous-jacent est un arbre de GW critique, de variance finie (possédant des moments exponentiels), conditionné par $\{X = n\}$, et les marques Y sont i.i.d., indépendantes de l'arbre sous-jacent. On note σ_ξ^2 et σ_Y^2 les variances respectives de la loi de reproduction et des marques (cette dernière variance peut être infinie). Nous conservons les notations R_n, V_n, r_n, v_n . On note

$$v = \frac{2}{\sigma_\xi} e$$

et r le processus défini par $r = \sigma_Y \bar{r}$ où \bar{r} est le processus qui, conditionnellement sachant v , est un processus gaussien centré satisfaisant

$$\text{cov}(\bar{r}(s), \bar{r}(t)) = \check{v}(s, t).$$

2.2.1 Convergence du tour

Le premier résultat est le suivant

Théorème 7 [9] *Si $\mathbb{E}(Y) = 0$ et $\mathbb{E}(Y^2) < +\infty$, alors les lois finies dimensionnelles de (r_n, v_n) convergent vers celles de (r, v) .*

Ce théorème est une conséquence du fait (connu) que v_n converge vers v , et du TCL.

Le théorème suivant précise les conditions de convergence faible.

Théorème 8 [9] *Supposons $\mathbb{E}Y = 0$. On a équivalence entre les trois points suivants :*

- (i) $\mathbb{P}(|Y| \geq y) = o(y^{-4})$,
- (ii) $r_n \xrightarrow{(d)} r$ dans $C[0, 1]$,
- (iii) $(r_n, v_n) \xrightarrow{(d)} (r, v)$ dans $C[0, 1] \times C[0, 1]$.

La bonne condition sur les moments est donc celle-ci : il faut et il suffit d'avoir des moments d'ordre strictement supérieur à 4 pour la convergence du tour normalisé et donc pour la convergence du serpent discret normalisé vers le serpent brownien.

Commentaires

- ┆ Pourquoi y^{-4} est-il critique pour la convergence de r_n vers r ?
- ┆ La convergence d'une marche aléatoire, d'incrémentes $(X_i)_i$ i.i.d. centrée de variance σ^2 , normalisée à la Donsker, nécessite l'existence de moments d'ordre 2.

On a alors $\mathbb{P}(|X_1| \geq y) = o(y^{-2})$. Si cette condition est remplie on a

$$p_n = \mathbb{P}\left(\sup_{i \in [1, n]} \frac{|X_i|}{\sigma\sqrt{n}} \geq a\right) = 1 - (1 - \mathbb{P}(|X_i| \geq a\sigma\sqrt{n}))^n$$

tend vers 0 pour tout $a > 0$; si $\mathbb{P}(|X_1| \geq y)$ n'est pas négligeable devant y^{-2} , alors p_n ne tend pas vers 0. L'incrément normalisé $\frac{|X_i|}{\sigma\sqrt{n}} \geq a$ entraîne que la marche normalisée fait un saut de hauteur supérieure à a (on parlera de *pic*) avec une probabilité ne tendant pas vers 0. Ceci n'est pas compatible avec la convergence vers le brownien, même pour la topologie de Skorohod.

Revenons aux marches branchantes :

– On voit que des moments d'ordre 2 seulement suffisent pour que sur une branche choisie dans l'arbre, on ait convergence vers un brownien tué; c'est en substance, ce que dit le théorème 7.

– Pour avoir convergence du tour r_n dans $C[0, 1]$, il faut que toutes les marques normalisées $y(u)/n^{1/4}$ tendent vers 0. Or, il y a n marques et cette fois la normalisation par $n^{1/4}$ va faire apparaître la condition (i) du théorème.

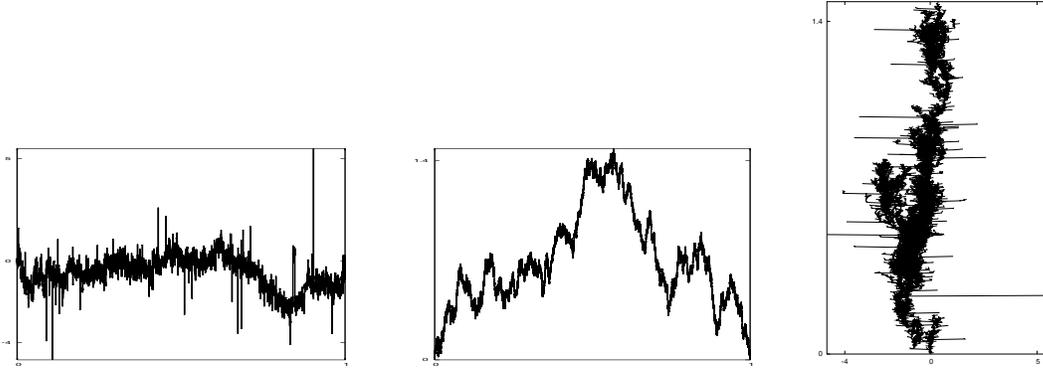


Figure 21: Simulation du tour normalisée (r_n, ν_n) d'une marche branchante avec $n = 10^5$, de marques de distribution symétrique satisfaisant à $\mathbb{P}(|Y| > y) = (1 + y)^{-5/2}$ pour $y \geq 0$. À droite, représentation du serpent correspondant

Le théorème suivant précise la convergence du tour :

Théorème 9 [9] (i) Si $\mathbb{E}Y = 0$ et $\mathbb{E}Y^2 < +\infty$ alors $r_n \xrightarrow{(d)} r$ dans $L^2[0, 1]$ et $(r_n, \nu_n) \xrightarrow{(d)} (r, \nu)$ dans $L^2[0, 1] \times C[0, 1]$.

(ii) Notons $\delta(x)$ la mesure de Dirac au point x . Supposons $\mathbb{E}Y = 0$ et $0 < \mathbb{E}Y^2 < \infty$, et soit

$$\nu_n = \frac{1}{n+1} \sum_{i=0}^n \delta\left(\sigma_Y^{-1} \sigma_\xi^{1/2} n^{-1/4} \Phi_{v_i}(h(v_i))\right) \quad (19)$$

la mesure empirique des points terminaux de la marche branchante. Alors $\nu_n \xrightarrow{(d)} \mu_{ISE}$ dans l'espace des mesures de probabilité sur \mathbb{R} où μ_{ISE} est définie par Aldous [21].

La mesure μ_{ISE} possède également une représentation selon le processus r (voir [51, 81, 3]) : si g est une fonction réelle, continue bornée, alors

$$\int g d\mu_{ISE} = \int_0^1 g(\sigma_\xi^{1/2} \sigma_Y^{-1} r(t)) dt. \quad (20)$$

Commentaires

Intuitivement les Théorèmes 7, 8 et 9 nous disent que lors des sauts du tour, seulement des petits sous-arbres sont déplacés comme on le voit sur la Figure 21. En effet, la mesure d'occupation du serpent discret est la même asymptotiquement que celle du serpent brownien, qui lui est continu (et il a aussi les mêmes finies dimensionnelles). Aldous [21] définit en fait μ_{ISE} comme étant la limite de ν_n (définie par la formule (19)). Dans (ii), on montre précisément l'affirmation d'Aldous : la suite de mesures ν_n converge vers une mesure (aléatoire) universelle. Celle-ci possède la représentation donnée dans (20).

LE TOUR CHEVELU :

Lorsque Y ne possède pas de moment d'ordre $4 + \varepsilon$, afin d'observer la convergence de r_n , nous devons changer de topologie. Nous avons choisi de montrer que le graphe du tour convergeait, en temps qu'ensemble fermé.

Nous procédons comme suit : soit \mathcal{K} l'ensemble des compacts non vides inclus dans $[0, 1] \times \mathbb{R}$, muni de la distance de Hausdorff

$$d(K_1, K_2) = \max \left(\max_{x \in K_1} d(x, K_2), \max_{x \in K_2} d(x, K_1) \right). \quad (21)$$

On identifie une fonction continue de $[0, 1]$ dans \mathbb{R} avec son graphe qui est un élément de \mathcal{K} . On a, pour $f_n, f \in C[0, 1]$, $f_n \rightarrow f$ dans $(C[0, 1], \|\cdot\|_\infty)$ sssi $f_n \rightarrow f$ dans \mathcal{K} .

Soit $f \in C[0, 1]$ et $\Xi \subset [0, 1] \times (\mathbb{R} \setminus \{0\})$ un ensemble tel que $\Xi \cap ([0, 1] \times (\mathbb{R} \setminus [-a, a]))$ est fini pour tout $a > 0$; on définit $H(f, \Xi)$ comme étant l'union du graphe de f avec les segments verticaux $[(x, f(x)), (x, f(x) + y)]$ pour $(x, y) \in \Xi$. C'est-à-dire que $H(f, \Xi)$ est le graphe de f auquel on a ajouté des cheveux verticaux, dont la position, la longueur et la direction (vers le haut, vers le bas) sont décrits par Ξ (voir Fig. 22). On peut vérifier sans problème que $H(f, \Xi)$ est compact et donc

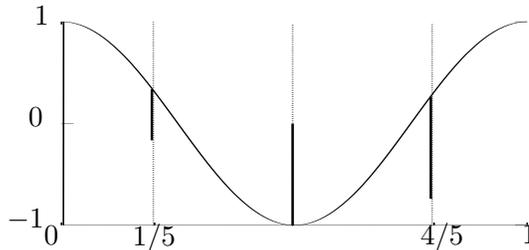


Figure 22: Représentation de $H(f, \Xi)$ avec $f(x) = \cos(2\pi x)$ et $\Xi = \{(\frac{1}{5}, -\frac{1}{2}), (\frac{1}{2}, 1), (\frac{4}{5}, -1)\}$

appartient à \mathcal{K} . De plus, l'application $(f, \Xi) \mapsto H(f, \Xi)$ est mesurable pour les tribus naturelles; nous allons maintenant considérer des f et des Ξ aléatoires.

Le cas où $\mathbb{P}(Y \geq y)$ est exactement de l'ordre de y^{-4} est le cas critique :

- nous n'avons pas convergence vers le serpent brownien à cause de ces pics dont nous avons parlé plus haut.
- Les pics possèdent un comportement asymptotique (sous des hypothèses de régularité de $\mathbb{P}(Y \geq y)$) qui sont du même ordre qu'une certaine contribution continue.

CAS CRITIQUE : "ÉQUILIBRE" ENTRE SAUTS ET CONTRIBUTIONS CONTINUES

Théorème 10 [9] *Supposons que $\mathbb{E}Y = 0$ et $\mathbb{P}(Y \geq y) = (a_+ + o(1))y^{-4}$ et $\mathbb{P}(Y \leq -y) = (a_- + o(1))y^{-4}$ pour des réels $a_+, a_- \geq 0$. Soit Ξ un processus de Poisson ponctuel dans $[0, 1] \times (\mathbb{R} \setminus \{0\})$ d'intensité $4a_+y^{-5}dx dy$ pour $y > 0$ et $4a_-|y|^{-5}dx dy$ pour $y < 0$, indépendant de (r, v) . Alors $r_n \xrightarrow{(d)} H(r, \Xi)$ dans \mathcal{K} et $(r_n, v_n) \xrightarrow{(d)} (H(r, \Xi), v)$ dans $\mathcal{K} \times C[0, 1]$.*

Dans ce cas, on parle de *tour chevelu*.

CAS OÙ LES SAUTS SONT PRÉPONDÉRANTS.

Dans le cas où $\mathbb{P}(Y \geq y)$ est exactement de l'ordre de y^{-q} avec $q < 4$, la partie à sauts devient prépondérante devant la partie continue. La normalisation que l'on applique à R_n pour que les sauts restent bornés, réduit à néant la contribution continue de R_n à la limite, et on obtient

Théorème 11 [9] *Supposons que $\mathbb{E}Y = 0$, $\mathbb{P}(Y \geq y) = (a_+ + o(1))y^{-q}$ et $\mathbb{P}(Y \leq -y) = (a_- + o(1))y^{-q}$ pour un certain $q < 4$ et $a_+, a_- \geq 0$. Soit Ξ un processus de Poisson ponctuel dans $[0, 1] \times (\mathbb{R} \setminus \{0\})$ d'intensité $qa_+y^{-q-1}dx dy$ pour $y > 0$ et $qa_-|y|^{-q-1}dx dy$ pour $y < 0$, indépendant de v . Alors $n^{-1/q}R_n(2n\cdot) \xrightarrow{(d)} H(0, \Xi)$ dans \mathcal{K} , et ceci conjointement avec $v_n \xrightarrow{(d)} v$.*

Théorème 12 [9]

- Si $q < 2$, la conclusion du théorème 11 a lieu même si $\mathbb{E}Y \neq 0$.
- Si $q \leq 1$, la conclusion du théorème 11 a lieu même si $\mathbb{E}|Y| = +\infty$.

CAS BIAISÉ :

Dans le cas où la moyenne des marques est non-nulle, le processus R_n est, en “dehors des temps de saut”, quasi proportionnel à V_n . Les sauts sont du même ordre que la “contribution continue” dans le cas où $\mathbb{P}(Y \geq y)$ est de l'ordre de y^{-2} . Ceci est la conséquence du fait que pour “faire converger la contribution continue de R_n ”, on doit normaliser maintenant par $n^{1/2}$ le processus R_n .

Théorème 13 [9] *Supposons les hypothèses du Théorème 11 valables avec $q = 2$, et $\mathbb{E}Y = m$ éventuellement non nul, alors $n^{-1/2}(R_n(2n\cdot), V_n(2n\cdot))$ converge dans $\mathcal{K} \times C[0, 1]$ vers $(H(mv, \Xi), v)$ où Ξ est décrit dans le théorème 11.*

Ces sauts disparaissent à la limite si, moralement, les marques possèdent un moment d'ordre 2 :

Théorème 14 [9] *Si $\mathbb{P}(|Y| > y) = o(y^{-2})$, alors $n^{-1/2}(R_n(2n\cdot), V_n(2n\cdot)) \xrightarrow{(d)} (mv, v)$ dans $C[0, 1]^2$.*

2.2.2 Convergence des serpents

Sous les conditions du Théorème 8, grâce au théorème 6, le serpent discret converge vers le serpent brownien. Dans le cas où la limite du tour n'est pas continue, ce théorème n'est plus utilisable. Les deux théorèmes suivants ont pour but de décrire le comportement asymptotique des serpents discrets, lorsque leurs tours ne convergent pas dans $C[0, 1]^2$. Nous définissons une notion de serpent sauteur, et nous prouvons de nouveau la convergence du graphe des serpents normalisés.

Soit (f, ζ) un élément de \mathbb{T} . On rappelle que le serpent de tour (f, ζ) est l'élément (ν, ζ) de \mathcal{S} , défini par :

$$\nu(s, t) = f(\rho(s, t, \zeta)) \quad \text{pour } (s, t) \in [0, 1] \times [0, +\infty].$$

On définit le graphe du serpent de tour (f, ζ) par $\Gamma(f, \zeta)$:

$$\Gamma(f, \zeta) = \{(s, t, \nu(s, t)), (s, t) \in [0, 1] \times [0, +\infty]\}.$$

On pose \mathcal{K}' l'ensemble des compacts non-vides inclus dans $[0, 1] \times [0, +\infty] \times \mathbb{R}$ muni de la distance de Hausdorff. On autorise $t = +\infty$ de sorte que $\Gamma(f, \zeta)$ soit compact ; c'est une raison technique puisque $(s, t) = \nu(s, \zeta(s))$ pour $t \geq \zeta(s)$. La fonction $(f, \zeta) \mapsto \Gamma(f, \zeta)$ est continue de $C[0, 1]^2$ dans \mathcal{K}' (ceci joue le rôle du théorème d'homéomorphisme).

Soit $A \subset [0, 1] \times [0, +\infty] \times \mathbb{R}$, $\zeta \in C[0, 1]$, et $B \subset [0, 1] \times \mathbb{R}$ où $B \cap ([0, 1] \times (\mathbb{R} \setminus [-a, a]))$ est fini pour tout $a > 0$. On pose

$$(A, \zeta) \oplus B := \left\{ (x, y, z + a) : (x, y, z) \in A \text{ et } \begin{cases} a = 0 & \text{si } y < \zeta(x) \\ a \in [0, a^*] \text{ ou } [a^*, 0] & \text{si } y \geq \zeta(x), (x, a^*) \in B \end{cases} \right\}$$

L'ensemble (A, ζ) a pour vocation de représenter le graphe du serpent de durée de vie ζ et B celle de représenter les sauts du serpent ; l'ensemble $(A, \zeta) \oplus B$ a pour objet de décrire le graphe du serpent codé par (A, ζ) lors de l'ajout des segments codés dans B par leur position et leur longueur (voir commentaires suivants).

On définit le serpent sauteur $J(f, \zeta, X)$ avec ensemble de sauts $X \subset [0, 1] \times \mathbb{R}$, processus de durée de vie ζ , et processus des étiquettes continu f (compatible avec ζ), comme le compact

$$J(f, \zeta, X) = (\Gamma(f, \zeta), \zeta) \oplus X.$$

Théorème 15 [9] *Si les hypothèses du théorème 10 sont satisfaites, alors $\Gamma(r_n, v_n)$ converge faiblement dans \mathcal{K}' vers $J(r, v, \Xi)$ où Ξ est décrit dans le théorème 10.*

Une convergence similaire est obtenue lorsque les sauts écrasent la composante continue :

Théorème 16 [9] *Si les hypothèses du théorème 11 ou 12 sont satisfaites alors $\Gamma(n^{-1/q}R_n(2n\cdot), v_n)$ converge faiblement dans \mathcal{K}' vers $J(0, v, \Xi)$, où Ξ est décrit dans le théorème 11.*

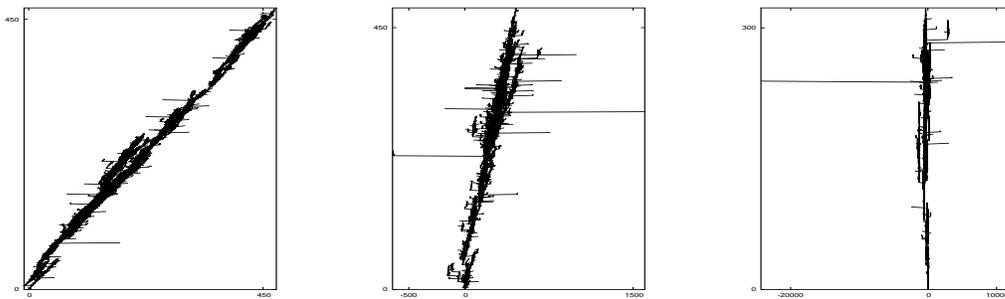


Figure 23: Simulation de marches branchantes (non normalisées) de taille 10^5 , de marque $1 + Y$ où Y est symétrique est satisfait à $\mathbb{P}(|Y| > y) = (1 + y)^{-5/2}$ pour la première, $\mathbb{P}(|Y| > y) = (1 + y)^{-3/2}$ pour la deuxième et $\mathbb{P}(|Y| > y) = (1 + y)^{-1.1}$ pour la dernière.

Commentaires

- Lorsque le graphe du tour du serpent discret converge vers un tour chevelu, comme dans le Théorème 10, le graphe du serpent discret correspondant, qui présente exactement les mêmes sauts que son tour, va converger vers un serpent à sauts. On a vu que les cheveux du tour limite étaient contenus dans les feuilles de l'arbre sous-jacent v (il n'y a pas de déplacement de sous-arbres). Ainsi, la limite du serpent discret – dans le cas où le tour présente asymptotiquement des sauts – va coïncider avec le graphe d'un serpent sans saut pour les points (s, t) où $(t < \zeta(s))$, puis, sur les feuilles $(s, \zeta(s))$ où un saut se produit, le serpent va présenter un saut instantané, égal au saut de r .

Sous les hypothèses du Théorème 15, la limite du graphe du serpent discret est le graphe d'un serpent brownien, sur lequel est greffé (moralement), à chaque endroit correspondant à un saut de r , un segment de longueur égale à ce saut. Le “moralement” provient du fait qu'une petite complication résulte du fait que le serpent est arrêté et non tué. Il s'agit donc plutôt d'une bande que d'un segment.

- Les techniques utilisées sont difficilement résumables. Comme en témoignent les théorèmes, il y a de nombreux phénomènes de compétition entre une partie “continue” et une partie “à sauts”. Le contrôle des contributions de chaque partie est faite à l'aide de techniques de troncatures.

2.3 Une propriété asymptotique de la correspondance par rotation

Le but de cette section est de présenter les résultats de [6]. Ces résultats sont en quelque sorte une application des résultats sur l'asymptotique des marches branchantes présentés plus haut.

Soit T_P^n l'ensemble des arbres planaires ayant n nœuds et T_B^n l'ensemble des arbres binaires ayant n nœuds internes, c'est-à-dire $2n + 1$ nœuds en tout. Il est bien connu que

$$\#T_P^{n+1} = \#T_B^n = \frac{(2n)!}{n!(n+1)!},$$

le n -ème nombre de Catalan. La correspondance par rotation fournit une bijection entre ces ensembles.

La correspondance par rotation Φ est plus généralement une application de l'ensemble des arbres planaires dans l'ensemble des arbres binaires. Elle peut être décrite via son action sur T_P^{n+1} pour $n \geq 0$:

$$\begin{aligned} \Phi : T_P^{n+1} &\longrightarrow T_B^n \\ \tau &\longmapsto \omega. \end{aligned}$$

L'arbre ω est construit comme suit (voir Fig. 24) : on efface la racine de τ . On conserve uniquement les arêtes [père - premier fils] et on ajoute une arête entre chaque nœud et son premier frère à sa droite (2ème étape de la figure 24). On fait "une rotation de $\pi/4$ " (3ème étape). Dans une 4ème étape, on ajoute des feuilles à cette structure interne.

NB : on peut écrire cette application sans problème dans le système de Neveu.

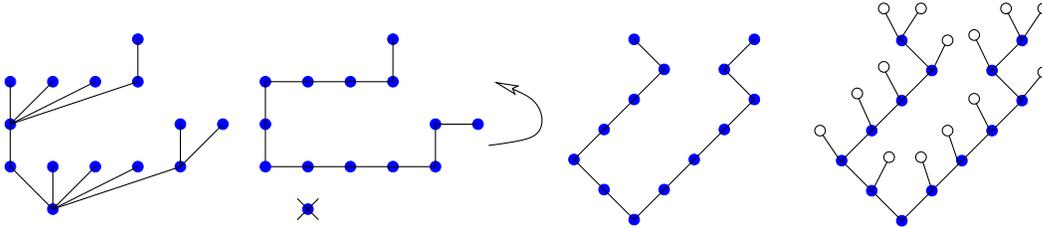


Figure 24: La correspondance par rotation

LA CORRESPONDANCE PAR ROTATION EST PRESQUE UNE HOMOTHÉTIE

Ce que l'on veut dire par là, c'est que, pour n grand, si on choisit un arbre planaire τ uniformément dans T_P^{n+1} , l'arbre binaire $\omega = \Phi(\tau)$ est presque égal à 2τ . Pour donner un sens à cet énoncé, on doit définir "2 τ " et "presque".

Supposons $T_P^{(n+1)}$ et $T_B^{(n)}$ munis tous deux de la loi uniforme. Soit τ un arbre planaire et $\omega = \Phi(\tau)$ l'arbre binaire associé. On note \mathcal{H}_n (resp. H_n) les processus des hauteurs de τ (resp. de ω), interpolés linéairement **et ramenés sur [0,1] en temps**. (Avant normalisation, le processus des hauteurs de τ a pour longueur n , celui de ω a pour longueur $2n$). D'après ce qui a été dit plus haut sur les arbres binaires et planaires, on a

$$\left(\frac{H_n(t)}{\sqrt{n}} \right)_{t \in [0,1]} \xrightarrow[n]{faible} (2^{3/2}e(t))_{t \in [0,1]} \quad \text{et} \quad \left(\frac{\mathcal{H}_n(t)}{\sqrt{n}} \right)_{t \in [0,1]} \xrightarrow[n]{faible} (2^{1/2}e(t))_{t \in [0,1]}. \quad (22)$$

Voyons maintenant les liens entre les deux processus \mathcal{H}_n et H_n .

Tout d'abord, regardons de plus près les arbres binaires. Soit v un nœud de l'arbre binaire ω . On note $r(v)$ (resp. $l(v)$) la profondeur à droite de v (resp. profondeur à gauche de v) : $r(v)$ est le

nombre de virages à droite en allant de la racine jusque v et $l(v)$ le nombre de virages à gauche, de sorte que $l(v) + r(v)$ est égal à $h(v)$ la profondeur de v ; on a $r(v) = a_{2,2}(v)$ et $l(v) = a_{1,2}(v)$ avec les notations de la page 22. D'autre part, il est clair que "monter d'un niveau dans τ " se traduit par "monter d'un pas vers la gauche" dans ω . Soit (v_0, v_1, \dots, v_n) la liste ordonnée des nœuds de τ et soit (w_0, \dots, w_{n-1}) la liste des nœuds internes de $\Phi(\tau)$. On a, pour tout $i \in \llbracket 0, n-1 \rrbracket$,

$$h(v_{i+1}) = l(w_i) + 1. \quad (23)$$

Pour ω dans T_B^n , on note R_n (resp. L_n) le processus des hauteurs à droite, (resp. processus des hauteurs à gauche) ramené sur $[0,1]$ en temps. On note H_n^I le processus des hauteurs de l'arbre composé des nœuds internes de l'arbre ω ramené sur $[0,1]$ en temps.

CONVERGENCE DES PROCESSUS DES HAUTEURS

Théorème 17 [6] *Chacun des processus*

$$\frac{2R_n}{\sqrt{n}}, \quad \frac{H_n^I}{\sqrt{n}}, \quad \frac{2L_n}{\sqrt{n}}, \quad \frac{H_n}{\sqrt{n}}, \quad \frac{2\mathcal{H}_n}{\sqrt{n}}$$

converge faiblement dans $(C[0,1], \|\cdot\|_\infty)$ vers $2^{3/2}e$. De plus, la différence entre n'importe quelle paire de ces processus tend vers 0.

Ceci implique, par exemple que

$$\left(\frac{2\mathcal{H}_n(t_1)}{\sqrt{n}}, \frac{H_n(t_2)}{\sqrt{n}} \right)_{(t_1, t_2) \in [0,1]^2} \xrightarrow[n]{faible} (2^{3/2}e(t_1), 2^{3/2}e(t_2))_{(t_1, t_2) \in [0,1]^2} \text{ dans } C[0,1]^2, \quad (24)$$

ce qui constitue un raffinement de (22). Le théorème 17 est une conséquence de résultats expliqués plus tôt dans ce document, en particulier du Théorème 2 et de (23).

L'homothétie et ISE

Soit u un arbre et c un réel positif. On définit cu comme l'arbre ayant la même structure arborescente que u mais de longueurs d'arêtes égales à celles de u multipliées par c (la longueur d'arête par défaut est 1).

Théorème 18 [6] *Asymptotiquement, la correspondance par rotation est une homothétie de rapport 2 : le couple d'arbres normalisés $(2^{1/2}n^{-1/2}\tau, 2^{-1/2}n^{-1/2}\omega)$ converge vers (T_∞, T_∞) où T_∞ est l'arbre aléatoire continu.*

Autrement dit, le couple des tours des deux arbres $2^{1/2}n^{-1/2}\tau$ et $2^{-1/2}n^{-1/2}\omega$ converge vers $(2e, 2e)$.

On peut en fait, estimer plus précisément la distance qui sépare ces deux arbres. Au passage, on montre un résultat concernant uniquement les arbres binaires.

Théorème 19 [6] *(il y a un serpent caché dans l'arbre binaire)*

La mesure empirique des différences entre le processus des hauteurs à droite et à gauche converge vers μ_{ISE} :

$$\frac{1}{2n+1} \sum_{j=0}^{2n} \delta \left(\frac{L_n(\frac{j}{2n}) - R_n(\frac{j}{2n})}{n^{1/4}} \right) \xrightarrow[n]{(d)} \mu_{ISE}. \quad (25)$$

De plus,

$$n^{-1/4} \|L_n - R_n\|_\infty \xrightarrow[n]{(d)} 2^{3/4}W \quad (26)$$

où W est le maximum de la valeur absolue du serpent brownien de durée de vie e .

Ce théorème et (23) permettent de montrer de manière élémentaire le théorème suivant, qui lui permet de préciser la distance entre les deux arbres τ et ω .

Théorème 20 [6] À propos de la différence entre \mathcal{H}_n et H_n^I :

$$n^{-1/4} \|2\mathcal{H}_n - H_n^I\|_\infty \xrightarrow{(d)} 2^{3/4}W. \quad (27)$$

De plus

$$\frac{1}{n} \sum_{k=0}^{n-1} \delta\left(\frac{2\mathcal{H}_n(\frac{k}{n}) - H_n^I(\frac{k}{n})}{n^{1/4}}\right) \xrightarrow{(d)} \mu_{ISE}. \quad (28)$$

Commentaires

- L'explication de l'apparition du serpent – et de sa mesure d'occupation ISE – est simple. On considère la marche branchante ayant pour arbre sous-jacent l'arbre binaire et ayant pour marque +1 pour les fils gauche, -1 pour les fils droits. Il est facile de montrer que la mesure d'occupation (soigneusement normalisée) de cette marche branchante converge vers μ_{ISE} et que son amplitude converge vers celle du serpent brownien (bien que les marques associées à chaque fratrie (+1, -1) ne vérifient pas les hypothèses demandées par les théorèmes présentés plus hauts, pour la convergence vers le serpent : les marques ne sont ni indépendantes, ni de même loi). D'autre part, la valeur terminale de la marche branchante sur un nœud u de l'arbre est simplement $l(u) - r(u)$.
- Peu de choses sont connues sur μ_{ISE} . La formule (25) donne une description de μ_{ISE} ne dépendant que d'un seul niveau d'aléa (un arbre binaire uniforme), alors que la description sous forme de marche branchante en nécessite deux. C'est à mon sens la plus simple description de μ_{ISE} sous forme de limite.

3 Des quadrangulations

Cette section est issue de [10], travail commun avec A. Mokkadem.

3.1 Modèles et problématique

Une carte planaire est un plongement propre – les arêtes ne se croisent que par leur sommet – d’un graphe connexe, non étiqueté, dans la sphère. Deux cartes sont considérées comme identiques si on peut envoyer l’une sur l’autre par un homéomorphisme direct de la sphère. Il existe une différence entre les notions de cartes planaires et de graphes planaires, puisqu’un graphe planaire est non plongé, et peut donc posséder plusieurs représentations non homéomorphes sur la sphère, comme sur la Figure 25.

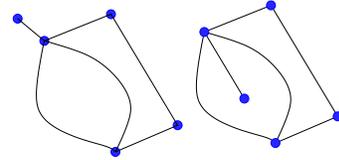


Figure 25: Deux cartes planaires différentes. La deuxième est une quadrangulation.

Une quadrangulation est une carte planaire dans laquelle toutes les faces sont de degré quatre (le degré d’une face est le nombre d’arêtes adjacentes à cette face, avec la convention que l’on compte deux fois une arête incluse dans une face); une quadrangulation est bipartie (c’est-à-dire les sommets sont 2-coloriables), sans boucle, et possède éventuellement des arêtes multiples. Une quadrangulation ayant n faces possède $2n$ arêtes et, par la formule d’Euler a $n + 2$ sommets.

On appelle quadrangulation pointée une quadrangulation dans laquelle un nœud est distingué et quadrangulation enracinée une quadrangulation dans laquelle une arête orientée est distinguée (cette fois on demande à l’homéomorphisme d’envoyer en plus la première marque sur la deuxième). On note $\vec{\mathcal{Q}}_n$ (resp. \mathcal{Q}_n^\bullet) l’ensemble des quadrangulations enracinées (resp. pointées) à n faces. Chaque quadrangulation enracinée possède une représentation “canonique” dans le plan obtenue par projection stéréographique : c’est la projection telle que l’arête racine (obtenue par projection) est adjacente à la face infinie du plan et telle que la face infinie soit à droite de l’arête racine, comme sur la Figure 26. Une fois représentée canoniquement dans le plan, ce sont les homéomorphismes du plan qui identifient les cartes.

PROBLÉMATIQUE : Nous nous intéressons ici à une question issue de la physique théorique, de la gravité quantique : peut-on définir une limite continue pour les quadrangulations ?

Des résultats concernant l’asymptotique de deux familles de cartes non normalisées ont été établis récemment. Les asymptotiques ont lieu pour une distance ultra-métrique : il est montré que, pour tout $k \in \mathbb{N}$, la restriction des cartes à un disque de rayon k autour de l’origine possède une limite faible. Une convergence faible des cartes vers une mesure aléatoire sur les cartes infinies est alors établie.

- (i) En 2003, Angel et Schramm [27] se sont intéressés aux triangulations ayant n faces, munies de la loi uniforme, et ont montré une limite faible vers une mesure aléatoire sur l’espace des triangulations infinies.
- (ii) En 2004, Chassaing et Durhuus [37] ont fait une étude du même type pour les quadrangulations ayant n faces, munies de la loi uniforme.

D’autre part, Chassaing et Schaeffer [41] ont montré que le diamètre des quadrangulations enracinées uniformes, normalisées par $n^{1/4}$, convergait en loi vers une variable (non dégénérée) liée au maximum du serpent brownien : ceci implique que seule une normalisation par $n^{1/4}$ peut permettre d’obtenir une limite continue, bornée, non triviale p.s.

Les travaux des physiciens à propos des cartes ont pour but de mener à bien certains calculs, d'établir certaines limites, dans le cadre d'espace métriques aléatoires. Il s'agirait de discrétiser certains modèles physiques, de faire certains calculs puis de passer à la limite. L'enjeu semble très important, mais malheureusement tout cela m'est encore très obscur. Quoiqu'il en soit, certains auteurs ont prédit que seule la normalisation par $n^{1/4}$ pouvait permettre d'obtenir une limite continue, non triviale et bornée. On pourra voir à ce propos les œuvres de Ambjørn, Durhuus, Watabiki, ... [26, 24, 25, 111].

MODÈLES PROBABILISTES CONSIDÉRÉS

- L'espace \mathcal{Q}_n^\bullet muni de la loi uniforme \mathbb{P}_U^n .
- L'espace $\vec{\mathcal{Q}}_n$ muni de la loi \mathbb{P}_D^n définie comme suit : soit q une quadrangulation dont le degré du nœud-racine est $\deg(q)$. On définit \mathbb{P}_D^n par :

$$\mathbb{P}_D^n(q) = \frac{1}{\deg(q)} \left(\sum_{q' \in \vec{\mathcal{Q}}_n} \frac{1}{\deg(q')} \right)^{-1}.$$

La loi \mathbb{P}_D^n peut également s'obtenir ainsi : on choisit une quadrangulation pointée sous \mathbb{P}_U^n , puis on choisit uniformément l'une des arêtes adjacentes au nœud racine et on la prend comme arête racine. La loi induite par cette procédure sur $\vec{\mathcal{Q}}_n$ est la loi \mathbb{P}_D^n .

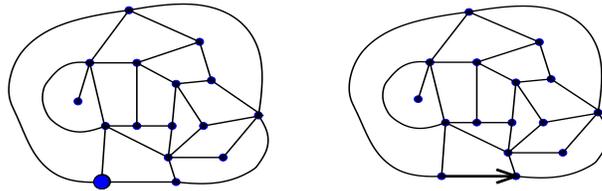


Figure 26: Une quadrangulation pointée et une quadrangulation enracinée

RÉSULTATS : Dans ce travail, nous construisons une notion de *carte abstraite*, comprenant aussi bien l'ensemble des quadrangulations normalisées qu'un modèle continu de cartes. Nous définissons une métrique sur l'ensemble des cartes abstraites. Nous prouvons ensuite que sous cette métrique, les quadrangulations normalisées convergent faiblement vers un carte continue aléatoire, que l'on baptise "carte brownienne".

Nous procédons selon les deux mêmes lignes directrices que pour l'obtention de l'arbre continu :

- Nous montrons qu'un certain codage des quadrangulations converge faiblement.
- Les cartes abstraites que nous construisons sont codées aussi bien par les codages discrets que limites. La distance que nous construisons sur l'ensemble des cartes abstraites est telle que la convergence des codages implique la convergence des cartes codées.

L'article [10] explicitant notre démarche est particulièrement long et dense ; je me contenterai d'indiquer quelques pistes permettant de comprendre les résultats obtenus.

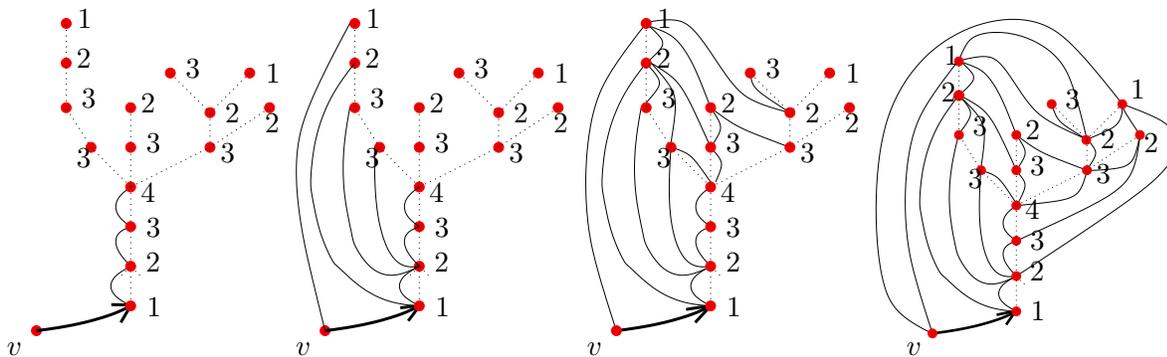


Figure 29: Construction de $Q(T)$ petit à petit.

quant au passage à l'infini.

CODAGE DES ARBRES BIEN ÉTIQUETÉS

Une quadrangulation se code de manière simple par un arbre bien étiqueté. Celui-ci peut se voir comme une marche branchante à n arêtes et dont les marques appartiennent à l'ensemble $\{+1, -1, 0\}$. On code donc les arbres bien étiquetés comme les marches branchantes, à l'aide des

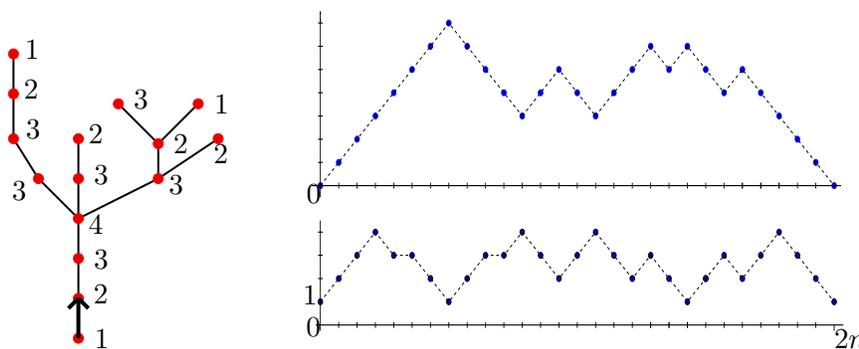


Figure 30: Codage des arbres bien étiquetés à la mode tour de serpent discret : au dessus V_n le tour de l'arbre sous-jacent, en dessous R_n , le processus des étiquettes.

deux processus compatibles V_n et R_n (voir Fig. 30). Ces deux processus codent la quadrangulation.

Noter bien que V_n est le tour d'un arbre : on appelle cet arbre, l'arbre glueur et on le note \mathcal{G}_n . Il s'agit en fait de l'arbre T dont on a effacé les étiquettes.

CONSTRUCTION DE LA QUADRANGULATION SUR LE CODAGE

Le deuxième point combinatoire fondamental dans ce travail est le suivant : on peut reconstruire directement la quadrangulation à l'aide du codage. Bien sûr, par composée des bijections, c'est évident. Mais cette reconstruction va se révéler particulièrement robuste pour le passage à la limite.

En partant de l'arbre bien étiqueté T , pour construire les arêtes de $Q(T)$, on a parcouru T en profondeur et on a ajouté une arête entre chaque coin d'étiquette k et le dernier coin précédemment visité d'étiquette $k - 1$. Or, le processus R_n contient les informations nécessaires pour ce travail : il code exactement une fois chaque coin de l'arbre T (sauf le coin sous la racine qui est codé 2 fois) et est porteur des étiquettes.

- On procède donc à la construction des arêtes directement sur R_n comme sur les Figures 31 et 32. Il apparaît alors que l'amas des arêtes construites par ce procédé constitue un arbre \mathcal{D}_n dont R_n est exactement le processus des hauteurs inverses (voir page 14). On appelle \mathcal{D}_n , l'arbre croulant.

• Les nœuds de la quadrangulation sont représentés éclatés sur R_n et sur V_n . On doit recoller les différents représentants d'un même nœud pour reconstruire $Q(T)$. Il y a plusieurs manières de procéder à ce recollage. L'une d'elle consiste à dessiner l'arbre \mathcal{D}_n directement sur l'arbre \mathcal{G}_n comme sur la Figure 33. Puis de recoller l'arbre \mathcal{G}_n (voir aussi commentaires suivants).

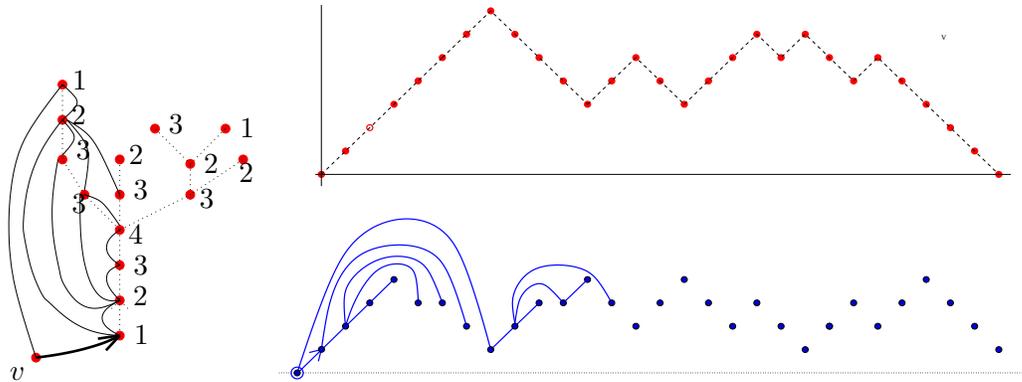


Figure 31: Début de la construction des arêtes ; en haut le processus V_n , en bas le processus R_n

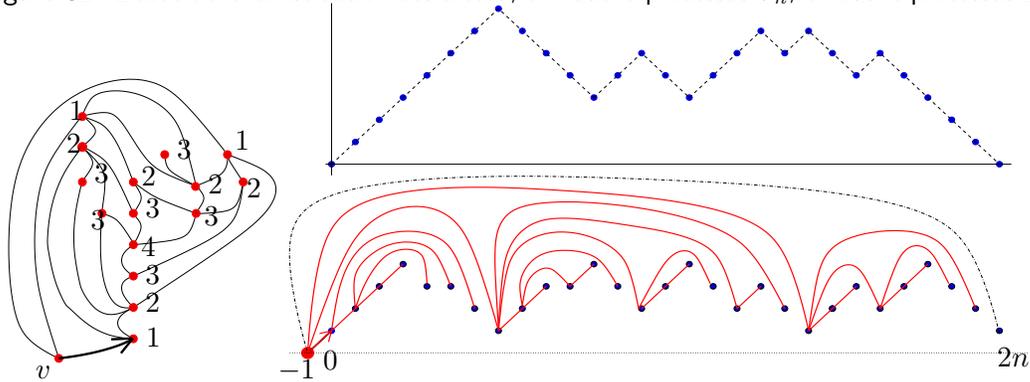


Figure 32: Fin de la construction des arêtes

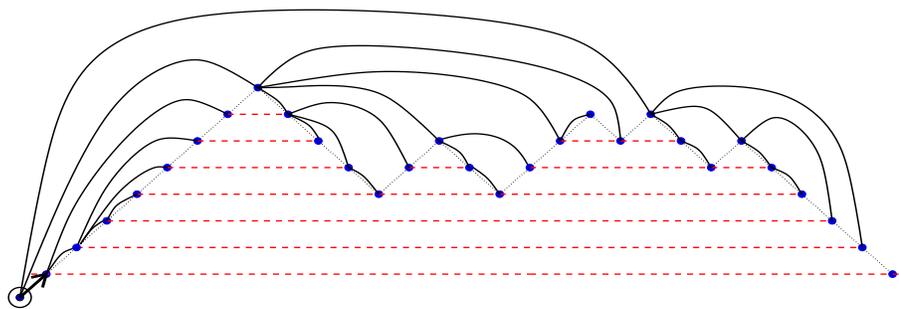
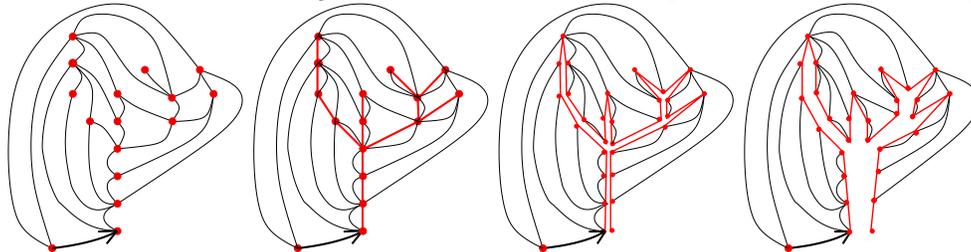


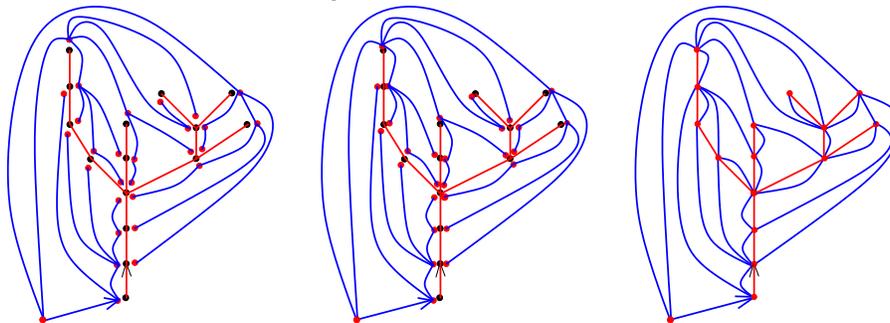
Figure 33: Construction de l'arbre \mathcal{D}_n directement sur les coins de \mathcal{G}_n : il reste juste à recoller l'arbre glueur pour obtenir $Q(T)$

Commentaires

En fait, il s'avère que si l'on “découpe” la quadrangulation selon l'arbre bien étiqueté, comme sur la série de figures suivantes, on s'aperçoit que les arêtes de la carte “s'organisent” sous forme d'un arbre : c'est l'arbre \mathcal{D}_n . Si l'on continue l'ouverture de l'arbre glueur on retrouve la représentation de la Figure 33.



Ainsi, pour recoller la carte, il suffit de mettre le k ème nœud de l'arbre \mathcal{D}_n (selon l'ordre inverse) en face du k ème coin de l'arbre \mathcal{G}_n (selon l'ordre direct) comme sur la série de figures ci-dessous et de recoller les nœuds de \mathcal{D}_n .



Cette vision de la carte : arbre croulant modulo les collages par l'arbre glueur va être imitée pour définir la notion de carte abstraite.

Nous normalisons le codage comme suit :

$$v_n(s) = \frac{V_n(2ns)}{\sqrt{n}} \quad \text{et} \quad r_n(s) = \frac{R_n(2ns) - 1}{n^{1/4}} \quad \text{pour } s \in [0, 1]. \quad (29)$$

Voici maintenant les grandes lignes de ce que l'on fait

[i] On montre que sous \mathbb{P}_D^n , le couple (r_n, v_n) converge faiblement (lorsque $n \rightarrow +\infty$). On note (r^+, v^+) sa limite.

[ii] Le couple (r_n, v_n) représente un couple d'arbres : r_n est le processus des hauteurs inverses de \mathcal{D}_n et v_n le tour (direct) de l'arbre glueur \mathcal{G}_n . La convergence de (r_n, v_n) vers (r^+, v^+) assure que le couple d'arbres $(\mathcal{D}_n, \mathcal{G}_n)$ converge vers un couple d'arbres limites $(\mathcal{D}_\infty, \mathcal{G}_\infty)$ dont le tour inverse du premier est r^+ et le tour du second est v^+ .

[iii] Obtenir $Q(T)$ à partir de $(\mathcal{D}_n, \mathcal{G}_n)$ se fait à l'aide d'un collage des deux arbres compatibles.

[iv] Nous allons imiter cette notion et coller les arbres limites (en fait, nous allons expliquer comment coller tout couple d'arbres “compatibles”).

Commentaires

| Le point [i] est vraiment difficile. Dans la Section 3.4 page 50, j'esquisse la démarche de la preuve.

3.3 Notion de carte “abstraite”

Nous avons construit un espace métrique \mathfrak{M} , ensemble des “cartes abstraites”, satisfaisant la liste de contraintes suivantes :

- Une quadrangulation enracinée discrète normalisée appartient à \mathfrak{M} .
- Chaque élément de \mathfrak{M} est un espace métrique (une carte) compact, connexe par arcs, paramétré, où prennent un sens les notions de nœuds, d’ordre cyclique autour des nœuds, de chemin, de cycle, d’intérieur de cycle.
- Chaque élément de \mathfrak{M} est codé à l’aide d’un couple d’arbres de la même manière que dans le cas discret.
- La convergence des codages des cartes implique la convergence des cartes associées.

Ceci a posé de nombreux problèmes, en particulier à cause du fait que les nœuds et les points des arêtes devaient absolument être distingués : on ne colle que les nœuds de \mathcal{D}_n . Ceci nous a conduit à définir une notion d’arbre (celle exposée Section 1.5) dont les objets “nœud” et “coin” faisaient partie intégrante.

Voici esquissée la notion de carte que nous avons définie. L’idée consiste à imiter le cas discret autant que possible.

Une carte “abstraite” est codée par un triplet $(\mathcal{D}, \mathcal{G}, b)$ où :

- $\star \mathcal{D} = (f, \mu_{\mathcal{D}})$ est un arbre dont f est le tour inverse (on l’appelle arbre croulant).
- $\star \mathcal{G} = (g, \mu_{\mathcal{G}})$ est un arbre dont g est le tour, on l’appelle arbre glueur.
- $\star b$ est une injection croissante de l’ensemble des nœuds de \mathcal{D} dans l’ensemble des coins de \mathcal{G} .

La carte M codée par $(\mathcal{D}, \mathcal{G}, b)$, se note $M = \text{MAP}(\mathcal{D}, \mathcal{G}, b)$. C’est un espace métrique compact dont voici les principales caractéristiques :

- Les points de M sont :
 - La racine de \mathcal{D}
 - Les points non-nœuds de \mathcal{D}
 - Les points de \mathcal{D} modulo les collages :

$$x \sim_b y \text{ ssi } b(x) = b(y).$$

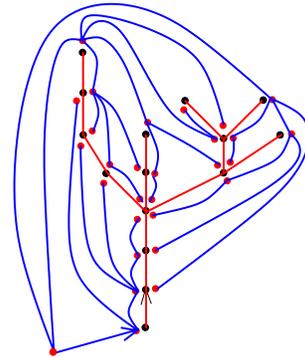
Ces derniers points sont appelés, avec la racine de \mathcal{D} , les nœuds de la carte M .

La carte M est connexe par arcs (car l’arbre croulant l’est). Elle est paramétrée par $[0, 1]$. En utilisant l’injection b , on peut définir autour de chaque point un ordre cyclique (infinitésimal). La métrique sur M que nous considérons est la suivante

$$d_M(\hat{u}, \hat{v}) = \inf_{k \geq 0} \inf \sum_{i=0}^k d_{\mathcal{D}}(u_{2i}, u_{2i+1}),$$

où $u_{2i+1} \sim_g u_{2i+2}$. Autrement dit, pour aller d’un point à un autre dans la carte, on se déplace sur les branches de l’arbre croulant entre u_{2i} et u_{2i+1} , puis on utilise un collage entre u_{2i+1} et u_{2i+2} , puis on marche sur l’arbre croulant, etc...

Nous définissons sur M des notions de chemin, de cycle, d’intérieur et d’extérieur de cycle, de degré des nœuds, etc....



Nous munissons l'espace des cartes \mathfrak{M} d'une métrique $d_{\mathfrak{M}}$ décrite schématiquement par :

$$d_{\mathfrak{M}}(M_1, M_2) = d_{\Gamma}(\mathcal{D}_1, \mathcal{D}_2) + d_{\Gamma}(\mathcal{G}_1, \mathcal{G}_2) + d(b_1, b_2).$$

La distance entre deux cartes tient compte de la distance entre les arbres utilisés et de la façon dont les injections b_i ont distribué les nœuds dans les coins.

- \mathfrak{M} contient les quadrangulations normalisées $M_n = \text{MAP}(\mathcal{D}_n, \mathcal{G}_n, b_n)$.
- \mathfrak{M} contient les cartes browniennes, définies par $M_{\infty} = \text{MAP}(\mathcal{D}_{\infty}, \mathcal{G}_{\infty}, b)$.

On a finalement tout construit pour pouvoir écrire le théorème suivant :

Théorème 21

$$M_n \xrightarrow[n]{\text{faible}} M_{\infty} \text{ dans } (\mathfrak{M}, d_{\mathfrak{M}})$$

3.4 Convergence de (r_n, v_n) sous \mathbb{P}_D^n

On a codé chaque quadrangulation de $\overrightarrow{\mathcal{Q}}_n$ par un couple (R_n, V_n) . Et comme on l'a vu plus haut, avoir une limite faible pour (R_n, V_n) permet d'obtenir une notion de limite pour les cartes. Mais ce point s'avère difficile. Que $\overrightarrow{\mathcal{Q}}_n$ soit muni de la loi uniforme ou de la loi \mathbb{P}_D^n , nous n'avons pas réussi à montrer directement la convergence faible du couple (R_n, V_n) normalisé sous la loi induite : nous n'avons même pas réussi à calculer les lois finies dimensionnelles dans le cas discret! Néanmoins, nous avons trouvé une preuve indirecte de cette convergence. Cette preuve me semble intéressante et au passage, elle permet de montrer la convergence des cartes pointées ayant n faces, normalisées, sous la loi uniforme. Voici les grandes lignes de cette preuve.

CODAGE ET QUADRANGULATIONS POINTÉES.

On appelle *arbre étiqueté* à n arêtes, un arbre planaire enraciné dont les nœuds portent une étiquette telle que : l'étiquette de la racine est 0 et la différence des étiquettes de deux nœuds adjacents appartient à $\{+1, -1, 0\}$. On a donc $\mathcal{W}_n^+ \subset \mathcal{W}_n$. Le tour des arbres étiquetés est beaucoup plus simple que celui des arbres bien étiquetés : sous la loi uniforme, le tour (R_n, V_n) des arbres étiquetés a même loi que le tour d'une marche branchante d'arbre sous-jacent un arbre planaire uniforme ayant n arêtes et de marques uniformes dans $\{-1, +1, 0\}$, indépendantes. Nous avons vu plus haut que sous ce modèle, le couple (r_n, v_n) normalisé converge vers le tour du serpent brownien. Nous allons nous servir de ce résultat.

Nous allons coder chaque quadrangulation de \mathcal{Q}_n^{\bullet} par une classe d'éléments de \mathcal{W}_n .

Nous définissons une opération de changement de racine sur les arbres étiquetés. Soit $T \in \mathcal{W}_n$. Pour $\theta \in \llbracket 0, 2n \rrbracket$, l'arbre $T^{(\theta)}$ est défini comme suit :

- $T^{(\theta)}$ et T sont égaux en tant qu'arbres non enracinés, non étiquetés.
- La racine de $T^{(\theta)}$ est obtenue en déplaçant la racine de T de θ arêtes (plus formellement, si F est la FPP de T , la racine de T^{θ} est $\overrightarrow{F(\theta)F(\theta+1)}$).
- Les étiquettes de $T^{(\theta)}$ s'obtiennent en ajoutant une constante à toutes les étiquettes de T de sorte que l'étiquette du nœud-racine de $T^{(\theta)}$ soit 1.

Le codage $(R_n^{(\theta)}, V_n^{(\theta)})$ de $T^{(\theta)}$ vérifie

$$V_n^{(\theta)}(i) = V_n(\theta \oplus i) + V_n(\theta) - 2\check{V}_n(\theta \oplus i, \theta) \quad \text{pour tout } i \in \llbracket 0, 2n \rrbracket, \quad (30)$$

$$R_n^{(\theta)}(i) = R_n(\theta \oplus i) - R_n(\theta) + 1 \quad \text{pour tout } i \in \llbracket 0, 2n \llbracket \quad (31)$$

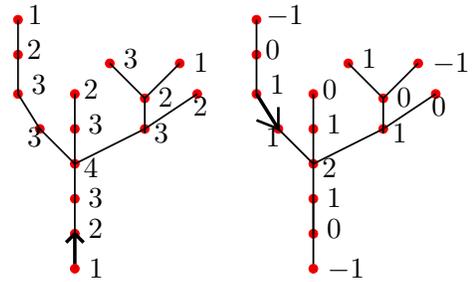


Figure 34: un arbre T et l'arbre $T^{(9)}$ obtenu en déplaçant de 9 arêtes l'arête racine de T

où le signe \oplus désigne l'addition modulo $2n$. Il est facile de voir que les opérateurs de changement de racine constituent un groupe isomorphe à $\mathbb{Z}/2n\mathbb{Z}$, on le note G_n . Deux arbres étiquetés T et T' de \mathcal{W}_n sont dits égaux modulo G_n , on note $T = T' \pmod{G_n}$, s'il existe $\theta \in \llbracket 0, 2n \rrbracket$ tels que $T^{(\theta)} = T'$.

ÉLÉMENT COMBINATOIRE DÉTERMINANT : Considérons une classe d'arbres étiquetés C modulo G_n : C contient des éléments positifs T_1, \dots, T_k , obtenus par les changements de racine sur les points d'étiquettes minimales de n'importe quel élément de C . Soit $Q(T_1), \dots, Q(T_k)$ les cartes enracinées obtenues par la bijection de Schaeffer. Il s'avère que les quadrangulations $Q(T_1), \dots, Q(T_k)$ vues comme quadrangulations pointées (en prenant comme racine leur nœud-racine) sont égales. Notons q^\bullet cette quadrangulation pointée. Si on considère tous les enracinages $\vec{q}_1, \dots, \vec{q}_l$ de q^\bullet (en prenant comme nœud racine le "point" et comme extrémité de l'arête racine un nœud adjacent à ce point), on s'aperçoit que $\{\vec{q}_1, \dots, \vec{q}_l\} = \{Q(T_1), \dots, Q(T_k)\}$. Enfin, puisque chaque quadrangulation pointée est enracinable, et que chaque quadrangulation enracinée appartient au moins à une classe C modulo G_n (d'après la bijection de Schaeffer) il existe une bijection entre \mathcal{Q}_n^\bullet et \mathcal{W}_n/G_n (ayant de nombreuses bonnes propriétés).

Ainsi, une quadrangulation pointée peut être codée par une classe d'arbres étiquetés, chacun d'entre eux étant codé par un tour de serpent discret. Nous allons coder les quadrangulations pointées par des classes de tour. Il restera à passer à la limite sur ces codages et à extraire les éléments positifs (en bijection avec les enracinages de la carte pointée).

PLONGEMENT DANS DES ESPACES FONCTIONNELS

Soit (R_n, V_n) le codage d'un élément T de \mathcal{W}_n et (r_n, v_n) la version normalisée comme dans (29). Le processus (r_n, v_n) appartient à l'espace du tour du serpent \mathbb{T} . Notons \mathbb{P}_S^n la loi de (r_n, v_n) lorsque T est choisi uniformément dans \mathcal{W}_n .

On considère le processus (r, v) à valeur dans \mathbb{T} dont la distribution est définie comme suit :

- $v \stackrel{(d)}{=} \sqrt{2}e$.
- Sachant v , le processus r est un processus gaussien centré, de fonction de covariance :

$$\text{cov}(r(s), r(t)) = \sqrt{2/3} \check{v}(s, t).$$

On note \mathbb{P}_S la loi de (r, v) . Il s'agit à un facteur près du tour du serpent brownien de durée de vie $\sqrt{2}e$; son support est dans \mathbb{T} . D'après le Théorème 5, on a

Proposition 1 *La suite de mesure $(\mathbb{P}_S^n)_n$ converge faiblement vers \mathbb{P}_S .*

Il nous reste à quotienter tout cela...

PLONGEMENTS ET CONVERGENCES

On définit une action de changement de racine sur un arbre quelconque (de tour de longueur 1) : soit $\mathcal{O} = \mathbb{R}/\mathbb{Z} \sim ([0, 1[, \oplus)$ où \oplus désigne maintenant l'addition modulo 1. Pour $(f, \zeta) \in \mathbb{T}$ et $\theta \in \mathcal{O}$, $(f^{(\theta)}, \zeta^{(\theta)})$ est défini par

$$\begin{aligned} f^{(\theta)}(x) &= f(\theta \oplus x) - f(\theta) && \text{pour tout } x \in [0, 1] \\ \zeta^{(\theta)}(x) &= \zeta(\theta \oplus x) + \zeta(\theta) - 2\check{\zeta}(\theta \oplus x, \theta) && \text{pour tout } x \in [0, 1]. \end{aligned} \quad (32)$$

Une action de $\mathcal{O} = ([0, 1[, \oplus)$ sur \mathbb{T} se définit aisément : pour $\theta \in \mathcal{O}$ le changement de racine J_θ est

$$\begin{aligned} J^{(\theta)} : \quad \mathbb{T} &\longrightarrow C[0, 1] \times C[0, 1] \\ (f, \zeta) &\longmapsto (f, \zeta)^{(\theta)} = (f^{(\theta)}, \zeta^{(\theta)}) \end{aligned}$$

où $(f^{(\theta)}, \zeta^{(\theta)})$ est défini en (32). Ceci correspond à un changement de racine sur l'arbre de tour ζ et par un changement d'origine dans le processus des étiquettes f .

Soit x dans \mathbb{T} ; notons $\pi(x)$ ou \bar{x} la classe de x . On note \mathbb{P}_S^n la loi de $\pi(r_n, v_n)$ lorsque (r_n, v_n) est de loi \mathbb{P}_S^n . Notons maintenant $\mathbb{P}_{\bar{S}}$ la distribution de $\pi(r, v) = \overline{(r, v)}$ lorsque (r, v) est de loi \mathbb{P}_S . Comme π est continue (si \mathbb{T}/\mathcal{O} est munie de la topologie quotient), la Proposition 1 implique

Théorème 22 *La convergence faible suivante a lieu*

$$\pi(r_n, v_n) \xrightarrow[n]{faible} \pi(r, v) \quad i.e. \quad \mathbb{P}_S^n \xrightarrow[n]{faible} \mathbb{P}_{\bar{S}}.$$

On doit maintenant raccrocher tout cela au modèle $(\mathcal{Q}_n^\bullet, \mathbb{P}_U^n)$. La loi \mathbb{P}_U^n induit la loi uniforme sur les codages discrets de \mathbb{T}/\mathcal{O} par la bijection $\mathcal{Q}_n^\bullet \sim \mathcal{W}_n/G_n$. Or \mathbb{P}_S^n n'est pas la loi uniforme car les classes modulo G_n n'ont pas toutes même cardinal (certains arbres possèdent des symétries). Mais on peut prouver que la distance en variation totale entre \mathbb{P}_U^n et \mathbb{P}_S^n tend vers 0. Ainsi, munir les quadrangulations pointées de la loi uniforme ou de la loi \mathbb{P}_S^n est équivalent en ce qui concerne les limites faibles associées.

Le Théorème 22 implique donc que le codage des quadrangulations pointées sous la loi uniforme converge faiblement. On a obtenu la convergence du codage des quadrangulations pointées. Comme il a été dit plus haut, enraciner uniformément une quadrangulation pointée sous \mathbb{P}_U^n induit la loi \mathbb{P}_D^n sur $\bar{\mathcal{Q}}_n$. Par la bijection $\mathcal{Q}_n^\bullet \sim \mathcal{W}_n/G_n$, ceci se traduit par : prendre un représentant positif uniformément dans une classe \mathcal{W}_n/G_n donne un élément (r_n, v_n) sous la loi induite par \mathbb{P}_D^n . Or, l'extraction d'un élément positif dans les classes de \mathbb{T}/\mathcal{O} est une opération qui passe bien à la limite grâce à un résultat dû à Thomas Duquesne : p.s., sous la loi $\mathbb{P}_{\bar{S}}$, chaque classe $\pi(r, v)$ contient un unique représentant positif (c'est-à-dire, il existe p.s. un unique $\theta \in \mathcal{O}$ tel que $r^{(\theta)} \geq 0$, $v^{(\theta)}$ étant lui toujours positif). Notons (r^+, v^+) cet élément positif. On montre alors que (r_n, v_n) sous \mathbb{P}_D^n converge faiblement vers (r^+, v^+) .

AUTRES RÉSULTATS

La convergence du codage (r_n, v_n) sous \mathbb{P}_D^n entraîne la convergence de nombreuses fonctionnelles des quadrangulations vers leurs versions limites. En particulier, le rayon (la distance maximum avec la racine), le profil (nombre de nœuds à distance successive de l'origine), etc...

QUADRANGULATIONS AVEC DES ARÊTES DE LONGUEUR ALÉATOIRE

Supposons que les arêtes possèdent une longueur aléatoire (indépendantes des autres arêtes et de la quadrangulation sous-jacente). Nous montrons que sous l'hypothèse d'existence de moments d'ordre $5 + \varepsilon$ pour ces longueurs, les quadrangulations convergent également vers la quadrangulation brownienne.

CONCLUSION

Il existe de très nombreux travaux combinatoires concernant les cartes (Tutte, Liskovets, l'école bordelaise de combinatoire dont Cori, Vauquelin, Schaeffer, Bousquet-Mélou, ...) : le but est de les décomposer, les compter, les mettre en bijection avec des objets plus simples. Récemment, Bouttier & al. [36] ont montré que de nombreux modèles de cartes discrètes possèdent une représentation sous la forme d'un arbre étiqueté que l'on doit recoller (et aussi partiellement effacer). Les travaux de Schaeffer et de Bouttier & al. suggèrent très fortement que, de la même manière que l'arbre continu est une limite quasi-universelle pour les arbres de GW conditionnés par la taille, la carte brownienne doit être la limite faible de nombreux modèles de cartes. Aujourd'hui, il me semble que la difficulté du passage à la limite tient finalement plus au problème de la convergence du codage qu'à des problèmes combinatoires.

II. Autres structures aléatoires et algorithmes

--*--*--*--*--*--

1 Parking avec densité

Les résultats de cette partie sont tirés de [1].

1.1 Données du problème

Soient m et n deux entiers vérifiant $0 < m \leq n$. Le problème de parking, introduit par Knuth [77] p. 545, est défini comme suit : m voitures cherchent à se garer sur un parking circulaire ayant n places numérotées $\{p_1, \dots, p_n\}$ selon la politique suivante :

- Les voitures $1, 2, \dots, m$ se présentent successivement dans le parking.
- La voiture numéro i **choisit** la place $p_{l(i)}$. Puis, elle se **gare** sur la première place libre parmi $p_{l(i)}, p_{l(i)+1}, p_{l(i)+2}, \dots$ avec la convention $p_{n+i} = p_i$.

Dans le modèle de Knuth, que l'on appellera *parking usuel*, les variables $(l(i))$ sont indépendantes et uniformes dans $\llbracket 1, n \rrbracket$. Ici, on considère un modèle dans lequel les voitures choisissent une place de manière non uniforme. On suppose que les variables $l(i)$ possèdent la représentation suivante :

$$l(i) = \lceil n X_i \rceil$$

où $(X_j)_{j \in \llbracket 1, m \rrbracket}$ est une liste de variables aléatoires i.i.d., admettant une densité f supportée par $[0, 1]$. On note $Par(n, m, f)$ le parking ainsi construit (voir Fig. 35). Le problème de parking usuel est donc simplement $Par(n, m, \mathbb{1}_{[0,1]})$, le parking "uniforme".

Notons $V(i)$ la place où se gare finalement la voiture i . Le déplacement de la voiture i , noté $Dep(i)$ est le nombre d'essais fructueux ou non que réalise la voiture i avant de se garer. On a par circularité du parking

$$Dep(i) = \begin{cases} V_i - l(i) + 1 & \text{si } l(i) \leq V(i) \\ n - l(i) + 1 + V(i) & \text{sinon.} \end{cases}$$

On appelle déplacement total dans $Par(n, m, f)$ la variable $d_{n,m}$ égale à la somme des déplacements :

$$d_{n,m} = \sum_{i=1}^m Dep(i).$$

Par exemple sur la Figure 35, on a $Dep(1) = 1, Dep(5) = 4$, et $d_{14,9} = 16$.

PROBLÉMATIQUE

Nous nous intéressons au comportement asymptotique de ces modèles de parking lorsque $n \rightarrow +\infty$ et pour $m = m(n)$. Voici les problèmes que l'on se pose.

Lorsque les m premières voitures sont garées :

- Quel est le déplacement de la $m + 1$ ème voiture dans le pire des cas ?
- Quelle est la loi, la moyenne, du déplacement total ?

La réponse à la première question est exactement 1 plus la taille du plus gros *bloc de voitures*, c'est-à-dire le nombre maximal de places consécutivement occupées après que les m premières voitures se sont garées (les tailles des blocs sur la Figure 35 sont 5, 2 et 2). La deuxième question dépend de l'évolution de la taille des blocs durant le remplissage du parking.

Dans ce travail, on étudie l'influence de f sur le comportement de ces variables. On montre que pour f non uniforme, il existe un phénomène de seuil pour $m = m(n)$ pour l'existence d'un bloc géant de l'ordre de n (dans le cas uniforme, il existe aussi un phénomène de seuil, mais il a lieu lorsque $m(n)/n \simeq 1$, c'est-à-dire, lorsque le parking est presque plein).

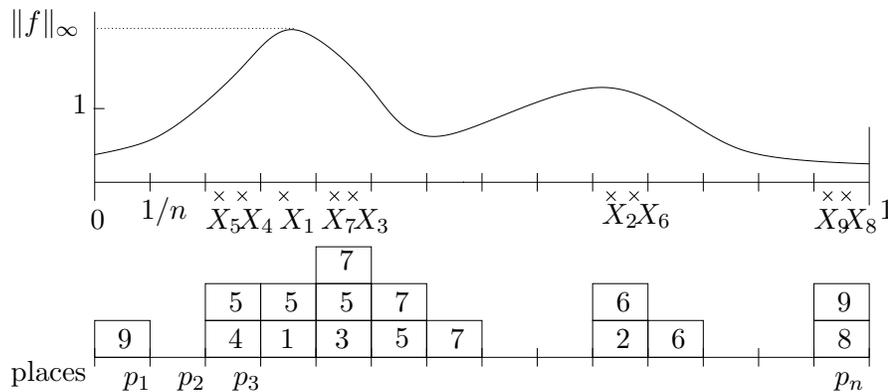


Figure 35: Sur ce dessin, on a représenté le choix de 9 voitures dans un parking ayant 14 places. D’abord, se gare la voiture 1, puis la 2, etc... On a représenté les essais des voitures sous forme de “couche”. Par exemple, la voiture 6 cherche à se garer au même endroit que la 2, puis à sa droite. La fonction f est la densité des variables X_i

Commentaires

Le problème de parking non-uniforme correspond en informatique au hachage linéaire où les données ne sont pas uniformes ; de telles hypothèses se trouvent dans les articles de Devroye [47], Gonnet [64], Pflug & Kessler [95] et Aldous [17].

L’ALGORITHME DE HACHAGE LINÉAIRE EN INFORMATIQUE : supposons que l’on ait des données x_1, \dots, x_m appartenant à un espace S et que l’on dispose de n cases mémoires. Supposons maintenant que l’on sache construire une fonction h (dite de hachage) de S dans $\llbracket 1, n \rrbracket$. Pour ranger les données dans les cases mémoires, on procède comme suit : on prend la donnée x_1 , on la met dans la case $h(x_1)$. On insère les autres données successivement. Pour $i \geq 2$, la donnée x_i est placée dans la première case libre parmi $h(x_i), h(x_i) + 1, h(x_i) + 2, \dots$ (modulo n). Si la fonction h est bien “mélangeante”, on va ranger rapidement les données. D’autre part, pour savoir si une donnée x est dans la table de hachage, on cherche à la place $h(x)$, puis $h(x) + 1$ etc... jusqu’au moment où on trouve la donnée ou une place vide. Afin d’étudier cet algorithme, on est amené à faire des hypothèses sur les valeurs $h(x_i)$: souvent on suppose que les données sont indépendantes et de même loi. Sous ces hypothèses, il est clair que le problème de parking et le hachage linéaire sont exactement le même problème. Ainsi, les caractéristiques : taille des blocs, déplacement total, etc... ont une interprétation immédiate en termes de coût de l’algorithme de hachage linéaire.

Pour le cas uniforme (c’est-à-dire lorsque $h(x_i)$ est uniforme dans $\llbracket 1, n \rrbracket$), on trouvera un état de l’art dans Chassaing & Louchard [39]. On pourra aussi consulter Janson [69, 68] et Pobleto & Viola [99] pour l’analyse d’autres politiques de hachage.

1.2 Résultats connus concernant le parking uniforme

Notons B_1, B_2, \dots la suite des largeurs des blocs ordonnés par ordre décroissant. On a

■ Parking plein : $m(n) = n$. On a $B_1 = n$ et

$$\frac{d_{n,n}}{n^{3/2}} \xrightarrow[n \rightarrow +\infty]{(d), Moments} \int_0^1 e(s) ds$$

Flajolet & al. [60] (1998). Voir [12], pour une preuve probabiliste.

■ Parking épars : $m(n) = \lfloor \alpha n \rfloor$ avec $0 < \alpha < 1$.

$$B_1 = \frac{1}{c(\alpha)} \left(\log n - \frac{3}{2} \log \log n + C_n \right) - 1$$

où C_n est une v.a. qui converge vers la variable doublement exponentielle (Pittel [98]) et

$$c(x) = x - 1 - \log x, \quad 0 < x < 1. \quad (33)$$

Le déplacement total possède une limite centrale [60] :

$$\frac{d_{n,m(n)} - \mu_{n,m(n)}}{\sigma_{n,m(n)}} \xrightarrow{(d)} \mathcal{N}(0,1) \quad \text{où}$$

$$\mu_{n,m(n)} = \mathbb{E}(d_{n,m(n)}) = \frac{\alpha}{2(1-\alpha)}m(n) - \frac{\alpha}{2(1-\alpha)^3} + O(n^{-1}), \quad (34)$$

$$\sigma_{n,m(n)}^2 = \text{Var}(d_{n,m(n)}) = \frac{6\alpha - 6\alpha^2 + 4\alpha^3 - \alpha^4}{12(1-\alpha)^4}m(n) - \frac{6\alpha^3 + 24\alpha^2 + 6\alpha}{12(1-\alpha)^6} + O(n^{-1}).$$

■ Changement de phase, parking presque plein : $m(n) = \lfloor n - \lambda\sqrt{n} \rfloor$ avec $\lambda > 0$. Chassaing & Louchard [39] étudient la dynamique de la taille des blocs, lorsque λ évolue. Ils montrent que

$$\frac{B_1}{n} \xrightarrow{(d)} B(\lambda)$$

où $B(\lambda)$ est la largeur de la plus grande excursion du processus

$$(\Psi_\lambda e)(x) = e(x) - \lambda(x) + \sup_{0 \leq y \leq x} \{\lambda y - e(y)\}.$$

Ils expriment également la limite jointe du k -uplet (B_1, B_2, \dots, B_k) en termes de variables gaussiennes ; ils montrent que les phénomènes de “coagulation de blocs” sont très fortement liés aux phénomènes de coagulations à noyau additif. D’autres travaux plus récents de Chassaing & Janson [38], Chassaing & Marchand [40], Bertoin [31], Bertoin & Miermont [33] montrent que les liens entre parking et coagulation (ou fragmentation) sont très riches.

1.3 Parking avec densité : résultats

On note $\|f\|_\infty = \text{ess sup } f$ (resp. $\text{ess inf } f$), le supremum essentiel de f (resp. infimum essentiel) de f , qui n’est pas supposée continue.

Théorème 23 [1] (*Phénomène de seuil*)

Soit $\alpha \in]0, 1[$ et soit $m(n) = \alpha n$. Dans le modèle $\text{Par}(n, \alpha n, f)$,

i) Cas sous-critique : si $\alpha \|f\|_\infty < 1$ alors

$$\frac{1}{c(\alpha \text{ inf } f)} \leq \liminf_n \frac{\mathbb{E}(B_1)}{\log n} \leq \limsup_n \frac{\mathbb{E}(B_1)}{\log n} \leq \frac{1}{c(\alpha \|f\|_\infty)}.$$

où la fonction c est définie par la formule (33).

ii) Cas sur-critique : si $\alpha \|f\|_\infty > 1$ alors il existe une constante $\gamma > 0$ telle que

$$\mathbb{P}(B_1 \geq \gamma n) \xrightarrow{n \rightarrow +\infty} 1.$$

Le cas (ii) montre que le cas uniforme est drastiquement différent puisque dans le cas uniforme le plus gros bloc est de l’ordre de $(\log n)/c(\alpha)$. Des renseignements beaucoup plus précis sur la taille des blocs et leur emplacement sont donnés dans [1].

CAS SOUS-CRITIQUE : $m(n) = \alpha n$ ET $\alpha \|f\|_\infty < 1$

Proposition 2 [1](Déplacement total dans les parkings épars)

Soit $\alpha \in]0, \|f\|_\infty[$; dans $\text{Par}(n, \alpha n, f)$, on a :

$$\frac{\mu_{n,(\alpha \inf)n}}{n} \leq \liminf_n \frac{\mathbb{E}(d_{n,\alpha n})}{n} \leq \limsup_n \frac{\mathbb{E}(d_{n,\alpha n})}{n} \leq \frac{\mu_{n,(\alpha \|f\|_\infty)n}}{n}$$

où $\mu_{n,m(n)}$ est défini par la formule (34).

L'argument donné ici repose sur un argument de couplage très simple. On compare notre modèle de parking à deux parkings uniformes ayant un nombre différent de voitures. Cet argument ne permet pas d'obtenir des résultats aussi fins que Pflug & Kessler [95] qui montrent, entre autres, que

$$\frac{\mathbb{E}(d_{n,\alpha n})}{n} \xrightarrow{n \rightarrow +\infty} \int_0^1 \frac{1 - (2\alpha - 1)f(x) + \alpha^2 f^2(x)}{2(1 - \alpha f(x))^2} dx.$$

CAS SURCRITIQUE : $m(n) = \alpha n$ AVEC $\alpha \|f\|_\infty > 1$

Cette fois, on analyse de plus près le *profil* du parking, défini comme suit : soit Y_k le nombre de voitures qui ont choisi la place p_k et H_k le nombre de voitures qui ont essayé de s'y garer (sur la Fig. 35 $H_1 = 1, H_2 = 0, H_3 = 2, H_4 = 2, \dots$ et $Y_1 = 0, Y_2 = 0, Y_3 = 2, Y_4 = 1, \dots$). La suite $(H_k)_{k=1, \dots, n}$ est appelée profil du parking (cette notion provient de Chassaing & Marckert [12]). Le profil contient de nombreuses informations relatives au parking, en particulier, elle contient la taille des blocs et aussi le déplacement total. Les suites $H.$ et $Y.$ sont liées par

$$H_{k+1} = (H_k - 1)_+ + Y_{k+1} \quad \text{pour tout } k \quad (35)$$

puisque toutes les voitures ayant essayé de se garer place p_k sauf une, essaient aussi place p_{k+1} . On peut alors facilement établir la formule suivante :

$$H_k = \max_{l \geq 0} \{ \dot{Y}_k + \dots + \dot{Y}_{k-l} - l \} \quad \text{pour tout } k \quad (36)$$

où, pour tenir compte du fait que le parking est circulaire, on a noté

$$\dot{Y}_k = Y_{k \bmod n} \text{ pour } k \neq 0 \text{ et } Y_0 = Y_n.$$

Il s'avère qu'à l'aide du théorème d'approximation forte de Komlos, Major & Tusnady [75]) il est possible de donner une limite p.s. du profil du parking, correctement normalisé. On note

$$\int_{a \curvearrowright b} f(x) dx = \begin{cases} \int_a^b f(x) dx & \text{si } 0 \leq a \leq b \\ \int_a^1 f(x) dx + \int_0^b f(x) dx & \text{si } 0 \leq b \leq a. \end{cases}$$

Théorème 24 [1] Soit h_n la fonction constante par morceaux, interpolant le profil $(H_k)_{k=0, \dots, n}$ de la manière suivante :

$$h_n(t) = \frac{H_{\lfloor nt \rfloor}}{m(n)}.$$

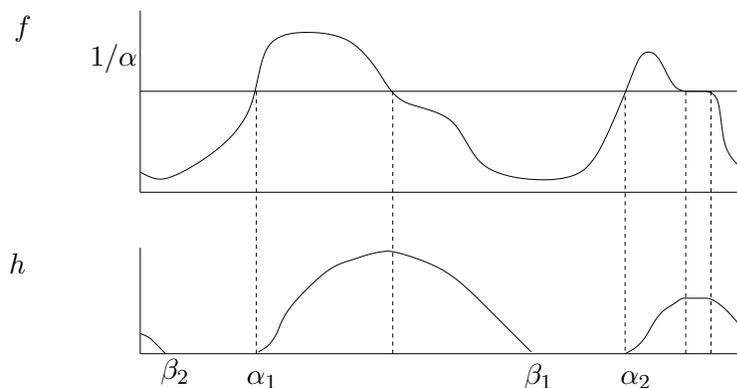
On a

$$h_n \xrightarrow[n \rightarrow +\infty]{p.s.} h$$

où $(h(t))_{t \in [0,1]}$ est la fonction déterministe

$$h(t) = \max_{x \geq 0} \left\{ \int_{(t-x) \curvearrowright t} f(u) du - \frac{x}{\alpha} \right\}.$$

Cette convergence p.s. a lieu pour la norme $\|\cdot\|_\infty$.

Figure 36: Dépendance schématique entre la densité et la limite h du profil

De cela, il s'ensuit que :

Corollaire 2 [1] • Soit $d_{m(n),n}$ le déplacement total dans $Par(n, \alpha n, f)$. On a

$$\frac{d_{m(n),n}}{\alpha n^2} \xrightarrow{p.s.} \int_0^1 h(t) dt.$$

• La place la plus visitée est visitée $\max H_k$ fois. On a

$$\frac{\max_{1 \leq k \leq n} H_k}{m(n)} = \max h_n \xrightarrow{p.s.} \max h.$$

Notons que h ne réalise pas son maximum là où f le réalise (dans le cas où f le réalise). Ainsi, dans le cas surcritique le déplacement total dans $Par(n, m(n), f)$ est de l'ordre de n^2 au lieu de $n^{3/2}$ dans le cas du parking uniforme plein. D'autre part, dans le cas du parking uniforme presque plein, l'analyse de Chassaing & Louchard montre que la place la plus visitée ne l'est que de l'ordre de $n^{1/2}$. Ici, elle est visitée de l'ordre de n fois.

Commentaires

D'autres résultats concernant la convergence de la taille et de l'emplacement des plus gros blocs sont également établis dans [1]; ils ne sont pas directement conséquence de la convergence du profil car la statistique "taille" ou "emplacement" des blocs n'est pas une fonctionnelle continue du profil normalisé.

2 Éléments sur la réduction de Lovász d'un réseau dans \mathbb{R}^n

Cette partie est tirée de [11], travail commun avec A. Akhavi et A. Rouault.

2.1 Introduction

Un réseau euclidien dans \mathbb{R}^n est l'ensemble des combinaisons linéaires entières de p vecteurs linéairement indépendants b_1, b_2, \dots, b_p dans \mathbb{R}^n (on supposera toujours $p \leq n$). L'espace \mathbb{R}^n est appelé espace ambiant. Un réseau peut être engendré par de nombreuses bases de cardinal p .

La réduction de réseau a pour but de trouver une base engendrant le réseau ayant de bonnes propriétés euclidiennes, à savoir : on cherche une base dont les vecteurs sont aussi "courts" et "aussi orthogonaux" que possibles. Il existe de nombreuses notions de réduction ; les deux plus célèbres sont dues à Minkowski et à Korkhine & Zolotarev. Trouver une base réduite au sens de Minkowski est un problème NP-complet. En pratique, on se contente de solutions approchées.

En 1982, Lenstra, Lenstra et Lovász [82] donnèrent un algorithme permettant de trouver une base "approximativement réduite" ; cet algorithme dépend d'un paramètre d'approximation $s \in]0, \sqrt{3}/2[$. On notera cet algorithme $\text{LLL}(s)$. L'algorithme $\text{LLL}(s)$ travaille par itérations successives et produit des bases de plus en plus réduites. Supposons qu'après k itérations, l'algorithme ait construit la base (b_1, \dots, b_p) . On procède à l'orthogonalisation de Gram-Schmidt des vecteurs b_i . On dit que la base (b_1, \dots, b_p) est $\text{LLL}(s)$ -réduite si pour tout $i \in \llbracket 1, p-1 \rrbracket$, le rapport des longueurs $\ell_{i,n}$ et $\ell_{i+1,n}$ des vecteurs orthogonalisés est minoré par s :

$$\ell_{i+1,n}/\ell_{i,n} \geq s \quad \text{pour tout } i \in \llbracket 1, p-1 \rrbracket.$$

Une borne supérieure pour le nombre d'itérations lorsque les vecteurs sont à coordonnées entières bornées est donné dans [82]. Daudé et Vallée [45] donnèrent quant à eux une borne supérieure pour la moyenne du nombre d'itérations dans le même cadre.

Dans ce travail, on considère des propriétés géométriques des réseaux euclidiens, sous le modèle uniforme, asymptotiquement lorsque la dimension de l'espace ambiant tend vers l'infini. Dans ce modèle, les p -vecteurs b_1, b_2, \dots, b_p , avec $p \leq n$, sont choisis de manière i.i.d., uniformément dans la boule euclidienne \mathcal{B}_n de \mathbb{R}^n . Nous montrons qu'une telle base possède des propriétés probabilistes asymptotiques quant au fait qu'elle soit directement, c'est-à-dire sans aucune itération de l'algorithme LLL , $\text{LLL}(s)$ -réduite.

Sous le modèle uniforme, la famille b_1, \dots, b_p est p.s. une famille indépendante au sens de l'algèbre linéaire. On appelle (b_1, \dots, b_p) une base aléatoire et le réseau engendré, un réseau aléatoire. L'entier

$$g := n - p$$

est appelé la *codimension* de la base b_1, \dots, b_p dans \mathbb{R}^n . La base est appelée *complète* si $g = 0$.

2.2 Résultats

Au début de ce travail, nous commençons par fournir des preuves probabilistes élémentaires à deux théorèmes "anciens" ; je les rappelle dans le paragraphe suivant car ils permettent de se faire une première idée quant aux caractéristiques asymptotiques des réseaux euclidiens aléatoires.

Théorème (Daudé & Vallée) *Soient $n \geq 1$, $p \in \llbracket 1, n \rrbracket$ et soit b_1, \dots, b_p une base aléatoire sous le modèle uniforme dans \mathbb{R}^n . Les carrés des longueurs $q_{1,n} = \ell_{1,n}^2, \dots, q_{p,n} = \ell_{p,n}^2$ des vecteurs orthogonalisés de Gram-Schmidt des vecteurs b_i sont indépendants et $q_{k,n}$ suit la distribution $\beta_{\frac{n-k+1}{2}, \frac{k+1}{2}}$.*

Théorème (Akhavi) Pour $i \geq 1$,

$$\frac{n}{2} (1 - q_{i,n}) \xrightarrow[n]{(d)} \gamma_{\frac{i+1}{2}}. \quad (37)$$

et

$$\frac{n}{2} q_{n-i,n} \xrightarrow[n]{(d)} \gamma_{\frac{i+1}{2}}. \quad (38)$$

Si $h(n) \rightarrow \infty$ et $h(n)/n \rightarrow 0$, alors

$$\frac{n}{h(n)} q_{n-h(n),n} \xrightarrow[n]{proba.} 1.$$

Si $0 < \alpha < 1$ et $k(n)/n \rightarrow 0$, on a

$$q_{\alpha n + k(n),n} \xrightarrow[n]{proba.} 1 - \alpha.$$

Les équations (38) et (37) indiquent que le comportement asymptotique des longueurs des vecteurs orthogonalisés dépend profondément de la valeur de i comparé à n . On va retrouver ce fait clairement dans les théorèmes qui suivent.

Définissons \mathcal{I}_n^g par

$$\mathcal{I}_n^g := \min_{1 \leq i \leq p-1} \frac{\ell_{i+1,n}^2}{\ell_{i,n}^2} = \min_{1 \leq i \leq n-(g+1)} \frac{\ell_{i+1,n}^2}{\ell_{i,n}^2}.$$

\mathcal{I}_n^g est le supremum de l'ensemble des s pour lesquels la base (complète ou incomplète) b_1, \dots, b_p est LLL(s^2)-réduite. On appelle cette quantité le *niveau de réduction de la base*.

Soit \mathcal{A}_n^g le plus petit indice i où le rapport $\ell_{n-i,n}^2/\ell_{n-i-1,n}^2$ vaut \mathcal{I}_n^g . On appelle \mathcal{A}_n^g l'*indice de pire réduction locale*.

Convergence du niveau de réduction

On définit le processus $(\mathcal{R}_j)_{j \geq 1}$ par

$$\mathcal{R}_j = \eta_j / \eta_{j+1} \quad \text{pour } j \geq 1$$

où les variables aléatoires η_i sont indépendantes, et $\eta_i \sim \gamma_{i/2}$. On définit pour tout entier K ,

$$\mathcal{I}_\infty^K := \inf \{ \mathcal{R}_j, j \geq K+1 \}.$$

Le théorème suivant donne le comportement asymptotique du niveau de réduction.

Théorème 25 [11] (i) Pour tout entier K fixé,

$$\mathcal{I}_n^K \xrightarrow[n]{(d)} \mathcal{I}_\infty^K$$

(ii) Soit $g : \mathbb{N} \rightarrow \mathbb{N}$ une fonction telle que $g(n) \leq n$ et $\lim_n g(n) = \infty$. Alors

$$\mathcal{I}_n^{g(n)} \xrightarrow[n]{proba.} 1.$$

La partie (i) du Théorème 25 pour $K = 0$ répond à une conjecture de Akhavi [16, 2 p.122].

UN MOT SUR LA PREUVE : Le processus $(q_{i,n}/q_{i+1,n})_{i=1,\dots,n-K-1}$ donne les rapports successifs des orthogonalisés. On veut montrer que l'infimum de ce processus "fini" converge vers l'infimum de $(\mathcal{R}_j)_{j \geq K+1}$, en loi. Notre démarche est la suivante : on construit des copies des processus $(q_{i,n}/q_{i+1,n})_i$ sur un espace de probabilité *ad hoc* et, après avoir prolongé les copies de manière convenable pour obtenir des processus à temps dans \mathbb{N} , on montre une convergence p.s. des copies vers un processus ayant même loi que \mathcal{R} dans un espace (de suites) l^p , pour un $p > 2$.

Quelques caractéristiques du niveau de réduction asymptotique \mathcal{I}_∞^K

Trouver la loi exacte de la variable \mathcal{I}_∞^K semble difficile ; toutefois, nous avons pu établir quelques résultats.

Proposition 3 [11] • *Pour tout K entier, la distribution de \mathcal{I}_∞^K a pour support $[0, 1]$ et admet une densité par rapport à la mesure de Lebesgue. De plus, on a*

(i)

$$P(\mathcal{I}_\infty^K \leq x) \underset{0+}{\sim} \frac{x^{(K+1)/2}}{\Gamma(\frac{K+2}{2})}.$$

(ii) *Il existe $\mu > 0$ tel que, pour tout $K \geq 0$, $P(\mathcal{I}_\infty^K \geq 1 - x) = O(e^{-\mu x^{-2}})$ dans un voisinage de 0.*

Convergence de l'indice de pire réduction locale

Ici, on donne des renseignements concernant l'asymptotique de l'indice i , où $\ell_{n-i+1,n}/\ell_{n-i,n}$ est minimum. Soit

$$\mathcal{A}_n^K = \min\{j \geq K, r_j^n = \mathcal{I}_n^K\} \quad , \quad K \in \mathbb{N}.$$

Proposition 4 [11] *Soit K entier fixé. Presque sûrement, il existe un unique entier \mathcal{A}_∞^K tel que $\mathcal{R}_{\mathcal{A}_\infty^K} = \mathcal{I}_\infty^K$. On a*

$$\mathcal{A}_n^K \xrightarrow[n]{(d)} \mathcal{A}_\infty^K \quad (39)$$

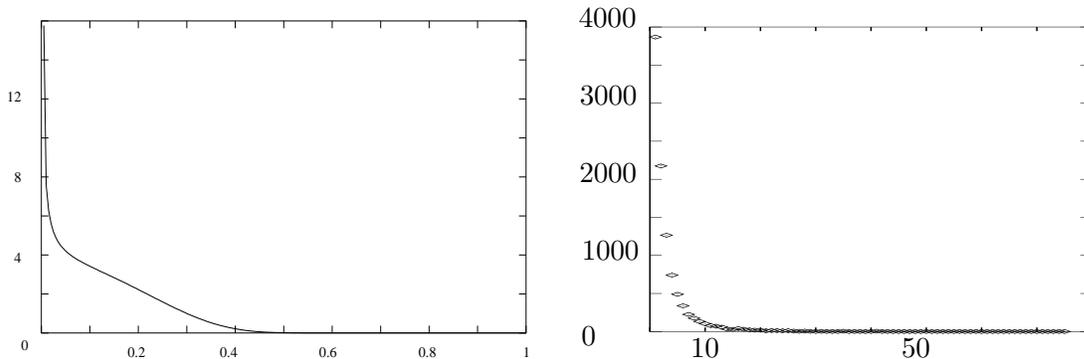


Figure 37: Simulation de la densité de \mathcal{I}_∞^0 à l'aide de 10^8 données, puis 10000 simulations de \mathcal{A}_∞

CONCLUSIONS

Lorsque g est un entier fixé, en particulier lorsque $g = 0$, \mathcal{I}_n^g converge en loi vers la variable \mathcal{I}_∞^g qui admet une densité et dont le support est tout l'intervalle $[0, 1]$; en conséquence, la probabilité qu'une base b_1, \dots, b_{n-g} soit s^2 -réduite converge vers une constante de $]0, 1[$ dépendant de s . Lorsque $g = g(n)$ tend vers $+\infty$, la probabilité qu'une base b_1, \dots, b_p soit réduite tend vers 1.

3 Arbres binaires de recherche et martingales

À propos du travail commun avec B. Chauvin, T. Klein et A. Rouault [8].

3.1 Arbres binaires de recherche et arbres de Yule

On appelle *arbre binaire* E , une partie de $\mathbb{V} = \emptyset \cup \bigcup_{n \geq 1} \{0, 1\}^n$, contenant \emptyset , contenant les préfixes des éléments de E et telle que $u0 \in E \Leftrightarrow u1 \in E$.

On introduit un objet que l'on appellera *arbre binaire de recherche étiqueté* (ABRE); il s'agit d'une structure très largement utilisée en informatique pour stocker des données appartenant à un espace A muni d'un ordre total (voir Mahmoud [89] et Devroye [46] pour un état de l'art).

On considère (x_1, \dots, x_n) une liste d'éléments différents de A , appelés *clefs*. L'ABRE associé à ces clefs est l'arbre binaire dont les nœuds internes sont étiquetés et sont construits comme suit : à chaque nœud interne est associée une (seule) clef appartenant à (x_1, \dots, x_n) . La clef x_1 est associée à la racine. Pour les autres clefs, on procède récursivement. Après les k premières insertions, on a un arbre ayant k nœuds, chacun porteur d'une des clefs de l'ensemble $\{x_1, \dots, x_k\}$. Pour insérer la $k + 1$ ème clef, on compare cette clef à celle de la racine de l'arbre. Si elle est inférieure, on recommence dans le sous-arbre gauche, sinon dans le sous-arbre droit. Si la racine de l'arbre ne possède pas encore de clef, on lui affecte la clef x_{k+1} (voir Fig. 38).

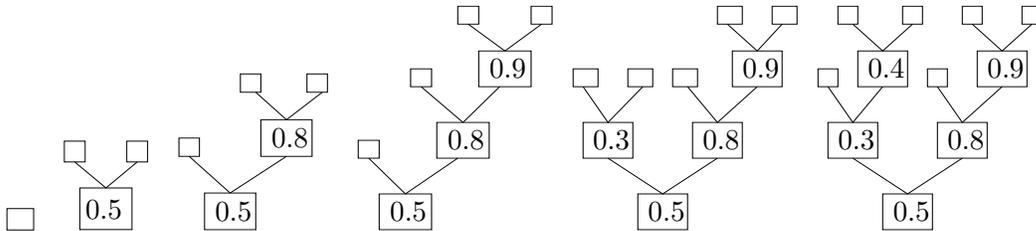


Figure 38: ABRE construit avec les clefs (0.5, 0.8, 0.9, 0.3, 0.4). Les carrés vides sont les feuilles qui sont ajoutés à la structure interne contenant les clefs.

On étudie le comportement de ces arbres lorsque les clefs sont des variables aléatoires i.i.d. uniformes sur $[0,1]$. L'ABRE obtenu est alors aléatoire; on le note L_n . L'arbre binaire sous-jacent, c'est-à-dire sans les étiquettes, sera appelé *arbre binaire de recherche* (ABR); on le note \mathcal{T}_n .

Sous la loi uniforme, (ou sous toute autre loi sans atome), la p.s. unique permutation $\sigma_n \in \mathcal{S}_n$ telle que

$$x_{\sigma_n(1)} < x_{\sigma_n(2)} < \dots < x_{\sigma_n(n)} \quad (40)$$

suit la loi uniforme, par échangeabilité des x_i . Il est bon de noter que \mathcal{T}_n ne dépend que de σ_n .

De plus, toujours par échangeabilité des x_i , σ_n est indépendant du vecteur réordonné $(x_{\sigma_n(i)})_{i=1, \dots, n}$. Sachant σ_n , l'insertion d'une nouvelle donnée x_{n+1} se fera dans l'un des $n + 1$ intervalles définis par (40) de manière uniforme. La signification en terme d'ABR est la suivante : sachant \mathcal{T}_n , la $n + 1$ clef a autant de chance de s'insérer dans chacune des $n + 1$ feuilles de l'arbre \mathcal{T}_n . Ainsi, la suite $(\mathcal{T}_n)_{n \geq 0}$ est une chaîne de Markov simple : au temps 0, elle est réduite à \emptyset . Au temps k , une des $k + 1$ feuilles est choisie avec la loi uniforme. On la transforme en nœud interne ayant 2 fils.

La chaîne L_n est beaucoup plus compliquée car elle dépend très fortement des valeurs insérées.

Commentaires

- Un algorithme de tri particulièrement efficace est l'algorithme dit *de tri rapide* (Quicksort). L'algorithme de tri rapide procède comme suit pour ordonner la suite (x_1, \dots, x_n) (de données différentes 2 à 2) :

Il fixe x_1 , dit le *pivot*, puis compare les éléments x_2, \dots, x_n à x_1 .

Il constitue un "paquet gauche" (resp. paquet droit) constitué des éléments inférieurs (resp. supérieur) à x_1 . Puis dans chacun des paquets, recommence la même procédure récursivement, jusqu'au moment où tous les paquets sont de taille 1 ou 0. À ce moment là, la liste est triée.

Il est facile de voir que l'ABR code l'exécution de quicksort. Stocker des données dans un ABR est vraiment efficace, et d'ailleurs, cette structure (ou des variantes) est implantée dans de nombreux logiciels. La raison tient essentiellement à (41), qui assure que l'on trouvera un élément dans la liste de données, ou qu'on saura qu'un élément n'y est pas, avec au plus $4.31... \log n$ comparaisons.

- La littérature concernant l'ABR et quicksort est absolument colossale. On pourra voir à ce sujet, les travaux de Devroye, Mahmoud, Rösler, Hwang, Neininger, Janson, Fill, Reed, Robson...

RÉSULTATS CONCERNANT LES ABR

On peut indiquer que la hauteur H_n et le niveau de saturation définis par

$$h_n = \min\{|u| : u \in \partial\mathcal{T}_n\} \quad , \quad H_n = \max\{|u| : u \in \partial\mathcal{T}_n\}$$

satisfont (Devroye [46]) ,

$$\lim_{n \rightarrow \infty} \frac{h_n}{\log n} \stackrel{p.s.}{=} c' = 0.3733... \quad \lim_{n \rightarrow \infty} \frac{H_n}{\log n} \stackrel{p.s.}{=} c = 4.31107... \quad (41)$$

où les constantes c' et c sont solutions de l'équation $\eta_2(x) = 1$ où

$$\eta_\lambda(x) := x \log \frac{x}{\lambda} - x + \lambda, \quad x \geq 0. \quad (42)$$

Sans aller plus loin, (41) montre que le comportement des ABR est drastiquement différent des arbres de GW sous le conditionnement $\{X = n\}$.

Nous avons étudié les ABR via un plongement dans un modèle continu en temps.

UN PROCESSUS DE FRAGMENTATION : on code les intervalles intervenant dans la description de la fragmentation ainsi : $I_\emptyset = (0, 1)$ et pour $u = u_1 u_2 \dots u_k \in \mathbb{V}$, on note I_u l'intervalle

$$I_u = \left(\sum_{j=1}^k u_j 2^{-j}, 2^{-k} + \sum_{j=1}^k u_j 2^{-j} \right).$$

Ainsi, chaque élément u de \mathbb{V} code un sous-intervalle I_u de $(0, 1)$ à extrémités dyadiques.

Le processus $(F(t))_{t \geq 0}$ est défini comme suit : $F(0) = I_\emptyset = (0, 1)$. Une variable $\tau_1 \sim \mathbf{Exp}(1)$ est associée à $F(0)$. Au temps τ_1 , le processus F saute : l'intervalle $(0, 1)$ se brise en son milieu et $F(\tau_1) = ((0, 1/2), (1/2, 1)) = (I_0, I_1)$. Après chaque temps de saut τ , les fragments de $F(\tau)$ se comportent indépendamment les uns des autres : chaque fragment I_u se brise en son milieu après un temps aléatoire de loi $\mathbf{Exp}(1)$, et crée I_{u0} et I_{u1} . Grâce à l'absence de mémoire des variables exponentielles, lorsque n fragments sont présents, chaque fragment a une probabilité de $1/n$ d'être le premier à se briser.

PROCESSUS DE YULE : On définit maintenant un processus à valeurs arbres binaires, à temps continu ; on le construit sur l'espace de la fragmentation. L'idée est simple ; elle consiste à interpréter les deux fragments I_{u0} et I_{u1} issus de I_u comme ses deux fils, I_{u0} (resp. I_{u1}) en étant le fils gauche

(resp. droit) (voir Fig. 39). Plus formellement, on définit l'arbre \mathbb{T}_t à l'aide de l'ensemble de ses feuilles $\partial\mathbb{T}_t$:

$$\partial\mathbb{T}_t = \{u, I_u \in F(t)\}.$$

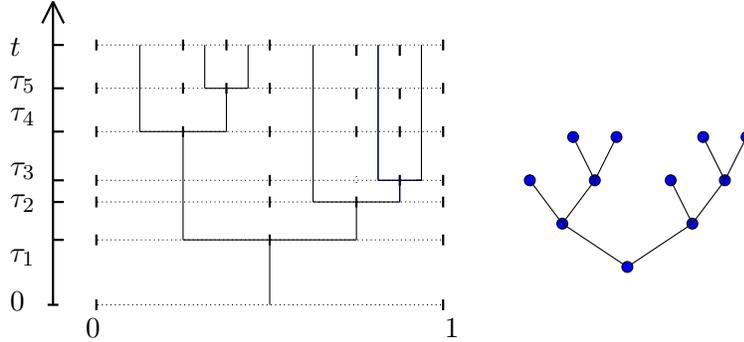


Figure 39: Construction de l'arbre de Yule à partir de la fragmentation.

On appelle processus de Yule, le processus à valeurs arbre binaire $(\mathbb{T}_t)_{t \geq 0}$. Le processus de comptage $(N_t)_{t \geq 0} := \#\partial\mathbb{T}_t$ donnant le nombre de feuilles dans \mathbb{T}_t est le processus de Yule classique (Athreya-Ney [29]). Notons que l'intervalle I_u présent dans la fragmentation est de longueur $2^{-|u|}$; ceci donne une correspondance simple entre la profondeur des feuilles dans l'arbre de Yule et largeur des intervalles dans la fragmentation.

Plongement des ABR dans l'espace de Yule

Les instants de sauts $0 = \tau_0 < \tau_1 < \tau_2 < \dots$ de \mathbb{T} . sont définis par

$$\tau_n = \inf\{t : N_t = n + 1\}. \quad (43)$$

Par construction les v.a. $(\tau_n - \tau_{n-1})_n$ sont indépendantes et indépendantes de $(\mathbb{T}_{\tau_n})_{n \geq 1}$. La variable $\tau_n - \tau_{n-1}$ est de loi $\mathbf{Exp}(n)$. Les deux processus $(\mathbb{T}_{\tau_n})_{n \geq 0}$ et $(\mathcal{T}_n)_{n \geq 0}$ ont même loi car il s'agit de deux processus de Markov avec même noyau de transition (ce fait était déjà connu, voir Tavaré [28] p.164 et Devroye [46]). Ainsi, sur l'espace de Yule (ou de la fragmentation), on peut construire en même temps l'arbre de Yule et l'ABR en posant simplement

$$(\mathbb{T}_{\tau_n})_{n \geq 0} = (\mathcal{T}_n)_{n \geq 0}. \quad (44)$$

Ce plongement va permettre d'établir des liens entre les martingales associées à ces deux familles d'arbres et déduire à partir des propriétés bien connues de l'arbre de Yule, de nouveaux résultats sur l'ABR.

DEUX MARTINGALES

Nous considérons deux filtrations sur l'espace des fragmentations : $(\mathcal{F}_t)_{t \geq 0}$ définie par $\mathcal{F}_t = \sigma(F(s), s \leq t)$ et $(\mathcal{F}_n)_{n \geq 0}$ définie par $\mathcal{F}_n = \sigma(F(\tau_1), \dots, F(\tau_n))$. Notons que la tribu $\mathcal{F}_n = \sigma(F(\tau_1), \dots, F(\tau_n))$ ne contient pas l'information des temps de saut, et contient "juste" la structure de l'arbre binaire ainsi que l'ordre de création des nœuds.

Un résultat classique (voir Uchuiyama [109]) est que $M(t, z)$ définie par

$$M(t, z) := \sum_{u \in \partial\mathbb{T}_t} z^{|u|} e^{t(1-2z)} \text{ pour } t \geq 0$$

est une \mathcal{F}_t martingale. Notons au passage que $M(t, 1/2) = 1$ et $M(t, 1) = e^{-t}N_t$.

D'autre part, notons

$$U_k(n) := \#\{u \in \partial\mathcal{T}_n, |u| = k\}, \quad k \geq 1,$$

le nombre de feuilles au niveau k dans \mathcal{T}_n . On appelle ici *profil* de \mathcal{T}_n , la suite $(U_k(n))_{k \geq 0}$.

Commentaires

D'un point de vue analyse d'algorithme, le vecteur $(U_k(n))_k$ est l'objet important. Pour placer une donnée dans l'arbre (ou pour chercher si elle se trouve), on compare cette donnée à celle de la racine, puis, on procède récursivement comme expliqué au début de cette section. On considère que le coût d'insertion d'une donnée est égal au nombre de comparaisons effectuées. Le profil de l'arbre fournit toutes les informations nécessaires pour analyser le coût (d'insertion d'une nouvelle donnée, de recherche d'une donnée ou le coût total de construction). Par exemple, une nouvelle donnée va s'insérer dans l'arbre au niveau k (et va nous coûter k donc) avec probabilité $U_k(n)/(n+1)$.

Pour obtenir des résultats asymptotiques sur le profil, nous le codons par ce que l'on nomme le *polynôme de niveau*, $\sum_k U_k(n)z^k$. On définit la suite de polynômes $(C_k)_k$ par

$$C_0(z) = 1 \text{ et pour } n \geq 1, \quad C_n(z) := \prod_{k=0}^{n-1} \frac{k+2z}{k+1} = (-1)^n \binom{-2z}{n}. \quad (45)$$

Pour $z \notin \frac{1}{2}\mathbb{Z}^- = \{0, -1/2, -1, -3/2, \dots\}$ et pour $n \geq 0$, on pose

$$\mathcal{M}_n(z) := \frac{1}{C_n(z)} \sum_{k \geq 0} U_k(n)z^k = \frac{1}{C_n(z)} \sum_{u \in \partial\mathcal{T}_n} z^{|u|}. \quad (46)$$

Jabbour & al. [42, 67] montrent que $(\mathcal{M}_n(z), \mathcal{F}_{(n)})_n$ est une martingale que l'on appellera martingale de l'ABR.

3.2 Liens entre les martingales

La formule de plongement (44) permet de lier les martingales $(\mathcal{M}_n, \mathcal{F}_{(n)})_n$ et $(M(t, z), \mathcal{F}_t)_t$. Il "suffit" d'extraire la composante d'espace de la martingale de Yule $M(\cdot, z)$, observée aux instants de sauts $(\tau_n)_n$. Soit $(\mathcal{E}_n(z))_n$ le processus \mathcal{F}_{τ_n} -adapté défini par

$$\mathcal{E}_n(z) := e^{\tau_n(1-2z)} C_n(z).$$

Un résultat (Athreya-Ney [29] ou Devroye [46] 5.4) affirme que, $e^{-t}N_t$ converge p.s. et que

$$\xi := \lim_{t \rightarrow \infty} e^{-t}N_t \sim \mathbf{Exp}(1). \quad (47)$$

Comme $\lim_n \tau_n = \infty$ p.s., de (43) et (47) on déduit que

$$\text{p.s. } \lim_n n e^{-\tau_n} = \xi.$$

Les martingales de Yule et de l'ABR sont ainsi liées par

Proposition 5 [8] *Soit $z \in \mathbb{C} \setminus \frac{1}{2}\mathbb{Z}^-$.*

1) *La famille $(\mathcal{E}_n(z))_{n \geq 0}$ est une martingale de moyenne 1, et*

$$\text{p.s. } \lim_n \mathcal{E}_n(z) = \frac{\xi^{2z-1}}{\Gamma(2z)}.$$

De plus, si $\Re z > 0$, la convergence a lieu dans L^1 .

2) Les deux martingales $(\mathcal{E}_n(z))_{n \geq 0}$ et $(\mathcal{M}_n(z))_{n \geq 0}$ sont indépendantes et

$$M(\tau_n, z) = \mathcal{E}_n(z) \mathcal{M}_n(z).$$

Le lien entre ces martingales et des martingales associées à d'autres modèles de structures arborescentes binaires est exploité dans Chauvin & Rouault [43].

PROPORTION DES NŒUDS DANS UN SOUS-ARBRE DONNÉ : Considérons $u \in \mathbb{V}$ et notons $\tau^{(u)} = \inf\{t : u \in \mathbb{T}_t\}$ le temps p.s. fini où u apparaît dans l'arbre de Yule ; pour $t > 0$, notons

$$\mathbb{T}_t^{(u)} = \{v \in \mathbb{V} : uv \in \mathbb{T}_{t+\tau^{(u)}}\}$$

le sous-arbre descendant de u et $N_t^{(u)} = \#\partial\mathbb{T}_t^{(u)}$ son nombre de feuilles. Pour $t > \tau^{(u)}$, le nombre de feuilles dans le sous-arbre enraciné en u est $n_t^{(u)} := N_{t-\tau^{(u)}}^{(u)}$. La propriété de branchement et (47) impliquent que, p.s., pour tout $u \in \mathbb{V}$

$$\lim_{t \rightarrow \infty} e^{-t} N_t^{(u)} = \xi_u \quad , \quad \lim_{t \rightarrow \infty} e^{-t} n_t^{(u)} = \xi_u e^{-\tau^{(u)}} \quad ,$$

où $\xi_u \sim \mathbf{Exp}(1)$. De plus, si u et v ne sont pas dans la même ligne de descente, les variables ξ_u et ξ_v sont indépendantes. Puisque pour $t > \tau^{(u)}$,

$$n_t^{(u)} = n_t^{(u0)} + n_t^{(u1)} \quad \text{et} \quad \tau^{(u0)} = \tau^{(u1)},$$

on obtient

$$\frac{n_t^{(u0)}}{n_t^{(u)}} \xrightarrow{p.s.} U^{(u0)} := \frac{\xi_{u0}}{\xi_{u0} + \xi_{u1}}, \quad \frac{n_t^{(u1)}}{n_t^{(u)}} \xrightarrow{p.s.} U^{(u1)} := 1 - U^{(u0)} = \frac{\xi_{u1}}{\xi_{u0} + \xi_{u1}}; \quad (48)$$

ceci permet d'associer à chaque nœud de \mathbb{V} une variable $\mathcal{U}([0, 1])$ représentant, la proportion asymptotique (qui existe donc p.s.) des nœuds descendants d'un nœud u comparée à la population descendant de son père. Posons, en particulier

$$U := U^{(0)} = \frac{\xi_0}{\xi_0 + \xi_1} \quad (49)$$

par itération, on peut montrer que la proportion des nœuds descendants de u s'exprime ainsi :

$$p.s. \lim_{t \rightarrow \infty} \frac{n_t^{(u)}}{N_t} = \prod_{v < u} U^{(v)}, \quad (50)$$

où les $(U^{(v)})_{v \in \mathbb{V}}$ sont indépendantes entre nœuds non-frères et vérifient $U_{u1} = 1 - U_{u0}$.

PROPRIÉTÉ SIMILAIRE DANS L'ABR : par plongement, l'ABR possède la même propriété (en temps discret). On peut aussi l'observer directement sur la version étiquetée, l'ABRE. Chaque nœud u représente un sous intervalle J_u de $[0, 1]$ (par exemple $J_\emptyset =]0, 1[$, $J_0 =]0, x_1[$ et $J_1 =]x_1, 1[$: toutes les clefs appartenant à l'intervalle $[0, x_1]$ vont aller s'insérer dans le sous-arbre enraciné en 0, celle appartenant à $[x_1, 1]$ dans celui enraciné en 1). D'autre part, puisque les clefs sont de loi uniforme, la première clé x_K tombant dans J_u va scinder celui-ci uniformément en J_{u0} et J_{u1} , c'est-à-dire

$$|J_{u0}| = U_u |J_u| \quad \text{et} \quad |J_{u1}| = (1 - U_u) |J_u|,$$

pour une certaine variable U_u , uniforme sur $[0, 1]$, indépendantes des autres variables U_u . Ainsi, $|J_u|$ possède la représentation $\prod_{v < u} U^{(v)}$; la loi des grands nombres dit que la proportion des variables $(x_i)_{K < i < n}$ qui appartiennent à J_u est asymptotiquement p.s. $\prod_{v < u} U^{(v)}$, lorsque $n \rightarrow +\infty$: ainsi, la proportion asymptotique de nœuds dans le sous-arbre enraciné en U_u a la représentation (50).

La différence par rapport au processus de Yule, est que dans l'ABRE, la variable U_u est connue dès qu'une clef s'est insérée dans le nœud u , alors que cette information n'est présente que dans la tribu de queue dans le processus de Yule.

CONVERGENCE DES MARTINGALES

Nous donnons maintenant les principaux résultats concernant le comportement asymptotique des martingales de Yule et de l'ABR, ainsi que leur lien.

Théorème 26 [8] *Pour $q \in]1, 2[$, posons $\mathcal{V}_q := \{z : \sup_t \mathbb{E}|M(t, z)|^q < \infty\}$. Alors*

$$\mathcal{V}_q = \{z : f(z, q) > 0\} \quad \text{où} \quad f(z, q) := 1 + q(2\Re z - 1) - 2|z|^q.$$

Posons $\mathcal{V} := \cup_{1 < q < 2} \mathcal{V}_q$, on a :

- a) Lorsque $t \rightarrow \infty$, $\{M(t, z)\}$ converge, p.s. et dans L^1 , uniformément sur tout compact C de \mathcal{V} .
- b) Lorsque $n \rightarrow \infty$, $\{\mathcal{M}_n(z)\}$ converge, p.s. et dans L^1 , uniformément sur tout compact C de \mathcal{V} .

Si $z > 0$, la martingale positive $\{\mathcal{M}_n(z)\}$ converge ps. vers $\mathcal{M}_\infty(z)$ qui est p.s. positive si $z \in (z_c^-, z_c^+)$, avec

$$z_c^- = c'/2 = 0.186\dots, \quad z_c^+ = c/2 = 2.155\dots$$

et $\mathcal{M}_\infty(z) = 0$ pour $z \notin [z_c^-, z_c^+]$ (Jabbour [67]). Le Théorème 26 donne une réponse à une question de [42] à propos du domaine complexe où la martingale $\{\mathcal{M}_n(z)\}$ est L^1 -convergente et uniformément convergente. Grâce au Théorème 27 II], ceci donne le domaine optimal de convergence L^1 sur \mathbb{R} .

Théorème 27 [8] I] *Soit $z \in (z_c^-, z_c^+)$.*

1) *La connection entre les limites des martingales est donnée par*

$$p.s. \quad M(\infty, z) = \frac{\xi^{2z-1}}{\Gamma(2z)} \mathcal{M}_\infty(z),$$

où ξ est défini en (47).

2) *Nous avons les formules de duplications suivantes :*

a) *Pour le processus de Yule,*

$$M(\infty, z) = z e^{(1-2z)\tau_1} (M_0(\infty, z) + M_1(\infty, z)) \quad p.s.$$

où $M_0(\infty, z)$ et $M_1(\infty, z)$ sont des copies indépendantes de $M(\infty, z)$, et sont indépendants de τ_1 .

b) *Pour l'ABR,*

$$\mathcal{M}_\infty(z) = z (U^{2z-1} \mathcal{M}_{\infty,(0)}(z) + (1-U)^{2z-1} \mathcal{M}_{\infty,(1)}(z))$$

où $\mathcal{M}_{\infty,(0)}(z)$ et $\mathcal{M}_{\infty,(1)}(z)$ sont des copies indépendantes de $\mathcal{M}_\infty(z)$, et sont indépendantes de U (qui est défini dans (49)).

II] *Enfin, hors de l'intervalle (z_c^-, z_c^+) , on n'a pas convergence L_1 . Pour $z \in (0, \infty) \setminus (z_c^-, z_c^+)$, on a p.s. $\lim_t M(t, z) = 0$ et $\lim_n \mathcal{M}_n(z) = 0$.*

MARTINGALES DÉRIVÉES

Les dérivées

$$M'(t, z) := \frac{d}{dz} M(t, z), \quad \mathcal{M}'_n(z) := \frac{d}{dz} \mathcal{M}_n(z)$$

sont également des martingales qui ne sont plus, cette fois, nécessairement positives. On a

Théorème 28 [8] *Pour $z \in (z_c^-, z_c^+)$, les martingales $(M'(t, z), t \geq 0)$ et $(\mathcal{M}'_n(z), n \geq 0)$ sont convergentes p.s.. Soient $M'(\infty, z)$ et $\mathcal{M}'_\infty(z)$ leur limite.*

a) *Nous avons la connection suivante :*

$$M'(\infty, z) = \frac{\xi^{2z-1}}{\Gamma(2z)} \left(\mathcal{M}'_\infty(z) + 2 \left(\log \xi - \frac{\Gamma'(2z)}{\Gamma(2z)} \right) \mathcal{M}_\infty(z) \right) \quad p.s.$$

où $\xi \sim \mathbf{Exp}(1)$ est défini dans (47) et est indépendant de $\mathcal{M}_\infty(z)$ et de $\mathcal{M}'_\infty(z)$.

b) *Nous obtenons une formule de duplication*

$$\begin{aligned} \mathcal{M}'_\infty(z) &= zU^{2z-1}\mathcal{M}'_{\infty,(0)}(z) + z(1-U)^{2z-1}\mathcal{M}'_{\infty,(1)}(z) \\ &+ 2z(U^{2z-1}\log U)\mathcal{M}_{\infty,(0)}(z) + 2z((1-U)^{2z-1}\log(1-U))\mathcal{M}_{\infty,(1)}(z) \\ &+ z^{-1}\mathcal{M}_\infty(z) \end{aligned} \quad (51)$$

où $U \sim \mathcal{U}([0, 1])$ est défini dans (49), et les v.a. $\mathcal{M}'_{\infty,(0)}(z)$ et $\mathcal{M}'_{\infty,(1)}(z)$ sont indépendants (et indépendant de U) et de même loi que $\mathcal{M}'_\infty(z)$.

On donne également le comportement, ainsi que des formules de duplications pour les martingales dérivées aux points z_c^\pm .

En prenant $z = 1$ dans (51), on obtient le corollaire suivant :

Corollaire 3 [8] *On a*

$$M'(\infty, 1) = \xi \left(\mathcal{M}'_\infty(1) + 2(\log \xi + \gamma - 1) \right) \quad p.s. ,$$

où γ est la constante d'Euler, et $\mathcal{M}'_\infty(1)$ satisfait la version p.s. de l'équation dite de Quicksort :

$$\mathcal{M}'_\infty(1) = U\mathcal{M}'_{\infty,(0)}(1) + (1-U)\mathcal{M}'_{\infty,(1)}(1) + 2U\log U + 2(1-U)\log(1-U) + 1,$$

où $\mathcal{M}'_{\infty,(0)}(1)$ et $\mathcal{M}'_{\infty,(1)}(1)$ sont des copies indépendantes de $\mathcal{M}'_\infty(1)$, indépendants de U .

Ceci est une version p.s. de l'équation dite de quicksort, sujet de nombreuses recherches ces dernières années (Fill, Janson, Devroye, Neininger, Rösler, Rüschemdorf [57, 58, 48, 90, 103, 102]).

3.3 Convergence du profil

La moyenne du profil est parfaitement connue: on a $\mathbb{E}(U_k(n)) = 2^k S_n^{(k)}/n!$ où $S_n^{(k)}$ est un nombre de Stirling de première espèce. Hwang [66] a montré que pour $\ell > 0$, lorsque $n \rightarrow \infty$ et $k \rightarrow \infty$ tel que $r = k/\log n \leq \ell$:

$$\mathbb{E} U_k(n) = \frac{(2 \log n)^k}{k! n \Gamma(r)} (1 + o(1)). \quad (52)$$

On en déduit que pour tout $\ell > 0$:

$$\mathbb{E} U_k(n) = \frac{n^{1-\eta_2(\frac{k}{\log n})}}{\Gamma(\frac{k}{\log n}) \sqrt{2\pi k}} (1 + o(1)), \quad (53)$$

où le $o(1)$ est uniforme pour $k/\log n \leq \ell$ et η_2 a été défini formule (42).

Une information de grande qualité sur le profil a été donnée par Jabbour & al. [42]:

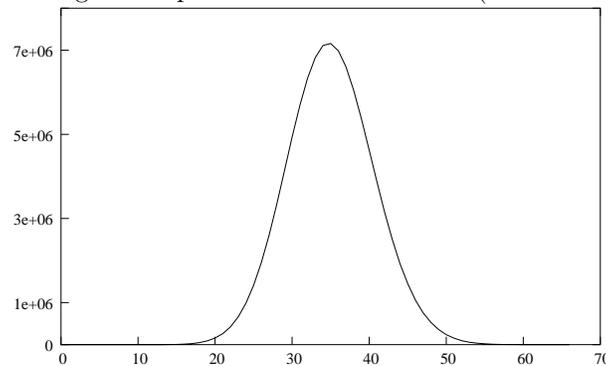
$$\text{p.s. } \lim_n \sup_{\frac{k}{\log n} \in [1.2, 2.8]} \left(\frac{U_k(n)}{\mathbb{E}U_k(n)} - \mathcal{M}_\infty\left(\frac{k}{2\log n}\right) \right) = 0. \quad (54)$$

Commentaires

Le profil est donc fortement lié à la martingale limite. Ceci montre que la famille de martingales $\mathcal{M}_n(z)$ n'est pas seulement un expédient technique.

En utilisant (54) et (53) on peut montrer que le profil de l'ABR, correctement normalisé, converge faiblement vers la courbe de Gauss (comme on le devine en voyant la figure suivante). Il suffit de prendre k de la forme $k = 2\log n + \lambda\sqrt{\log n}$ dans (54) et dans (53) et utiliser que $\mathcal{M}_\infty(k/(2\log n)) \rightarrow \mathcal{M}_\infty(1) = 1$.

Sur le dessin ci-dessous, on a simulé le nombre de feuilles sur les étages successifs d'un ABR de taille 10^8 ; il s'agit du profil d'un seul arbre (aucune moyenne n'a été prise).



On peut remarquer que pour déduire cette limite gaussienne, nous avons utilisé (54) seulement au voisinage de $z = 1$. Ainsi, (54) donne beaucoup plus d'informations que cela.

L'approche de Jabbour & al. reposait sur des estimations L^2 de $\mathcal{M}_n(z)$; ils pensaient que l'intervalle $[1.2, 2.8]$ pouvait être étendu à $I := (2 - 2^{1/2}, 2 + 2^{1/2}) = (0.585\dots, 3.414\dots)$, intervalle de z correspondant à la convergence L^2 de $\mathcal{M}_n(z)$.

Nous avons montré une extension de leur résultat à tous les compacts de la zone surcritique $(c', c) = (0.373\dots, 4.311\dots)$, en montrant dans chaque compact une convergence L^p , pour un p dépendant du compact.

Théorème 29 [8] *P.s., pour tout compact K de (c', c)*

$$\lim_n \sup_{k : (\frac{k}{\log n}) \in K} \left(\frac{U_k(n)}{E(U_k(n))} - \mathcal{M}_\infty\left(\frac{k}{2\log n}\right) \right) = 0.$$

L'utilisation de la martingale de Yule est encore déterminante dans la preuve de ce théorème. Les contrôles uniformes utilisent des résultats de Biggins [35] et Bertoin & Rouault [34].

III. Résultats issus de la thèse

○ ○ ○ ○ ○ ○ ○ ○ ○ ○

1 Couple hauteur-largeur des arbres binaires

Petit résumé de [13], travail commun avec P. Chassaing et M. Yor.

Soit E un arbre choisi uniformément parmi les arbres binaires ayant n nœuds internes. On note $W^{(n)}$ sa largeur, $H^{(n)}$ sa hauteur. On peut exprimer $W^{(n)}$ et $H^{(n)}$ à l'aide de la PPL S (pile du parcours en largeur) de E .

Sans trop de difficulté on peut montrer,

$$\frac{H^{(n)}}{\sqrt{2n}} \xrightarrow[n]{(d)} 2 \max_{0 \leq s \leq 1} e(s), \quad \frac{W^{(n)}}{\sqrt{2n}} \xrightarrow[n]{(d)} \max_{0 \leq s \leq 1} e(s),$$

Nous nous intéressons à la loi jointe et nous prouvons,

$$\left(\frac{H^{(n)}}{\sqrt{2n}}, \frac{W^{(n)}}{\sqrt{2n}} \right) \xrightarrow[n]{(d)} \left(\int_0^1 \frac{ds}{e(s)}, \max_{0 \leq s \leq 1} e(s) \right).$$

Nous donnons deux preuves de ce résultat.

ÉLÉMENTS DE LA PREMIÈRE PREUVE

Aldous [22, Th. 3] prouve que le couple de processus (tour, profil) des arbres de GW conditionnés par leur taille converge, vers $(2e, l/2)$, où l est le temps local au temps 1 de l'excursion e (il faut bien sûr normaliser). Dans le cas des arbres binaires, avec la bonne normalisation on obtient, en prenant le maximum

$$\frac{1}{\sqrt{2n}} (H^{(n)}, W^{(n)}) \xrightarrow{(d)} \left(2 \max_{0 \leq t \leq 1} e(t), \frac{1}{2} \max_{x \geq 0} l(x) \right).$$

En utilisant le résultat de Jeulin [78] qui dit que le temps local de l'excursion brownienne est une excursion changée de temps, on obtient une première preuve du résultat.

IDÉES DE LA DEUXIÈME PREUVE

La seconde preuve repose sur un raisonnement "discret". Deux points sont importants :

- La pile du parcours en largeur d'un arbre binaire ayant $2n+1$ nœuds est une excursion de Bernoulli de longueur $2n$ (au dernier pas près).

- le profil d'un arbre se lit sur la marche de Lukaciewicz (voir page 13) S_n . En particulier :

- Le maximum du profil, qui est égal à la largeur, vaut à peu près le maximum de la marche de Bernoulli (car les points sur lesquels on cherche le maximum sont écartés au plus de l'ordre de $n^{1/2}$ et donc les fluctuations entre ces points sont de l'ordre de $n^{1/4}$).
- La longueur du vecteur profil est égal à la hauteur. Or si k est l'abscisse d'un point du profil, le prochain point aura pour abscisse $k + S_k$. Ainsi, en tenant compte du fait que les variations de S_n sont relativement lentes et que le temps local en 0 est petit, on peut dire que moralement, j est l'abscisse d'un des points du profil avec "probabilité" $1/S_j$. Il s'ensuit, que la longueur du vecteur profil est à peu près $\sum_{j \geq 1} \frac{1}{S_j}$, puis, en passant à la limite $\int_0^1 \frac{1}{e_s} ds$.

2 Largeur des arbres de Cayley et parkings

Petit résumé de [12], travail commun avec P. Chassaing.

Les arbres de Cayley sont des arbres non ordonnés, dont les nœuds portent des étiquettes (ce sont des graphes connexes acycliques). Ils possèdent une représentation canonique sous la forme d'un arbre ordonné (voir Spencer [106]). Cet arbre (sans les étiquettes) a la forme d'un arbre à progéniture poissonnienne conditionnée par la taille.

Notons W_n la largeur d'un arbre de Cayley ayant n nœuds. On a

$$\frac{W_n}{\sqrt{n}} \xrightarrow[n]{(d)} \max\{e(s), s \in [0, 1]\}.$$

Ceci est une conséquence de la convergence du profil des arbres de GW conditionnés vers le temps local de l'excursion brownienne (voir Drmota & Gittenberger [49] et Pitman [96]) et du résultat de Jeulin [78]; ceci peut aussi se montrer en utilisant le fait que la différence entre le maximum du profil et de la PPL est petit.

Une question ouverte d'Odlyzko & Wilf [93] était la suivante : a-t-on également convergence des moments ?

Nous avons répondu positivement à cette question en donnant aussi une borne sur la vitesse de convergence. L'idée a été de construire les arbres de Cayley, où plus précisément la PPL à l'aide du processus empirique associé à n variables indépendantes, uniformes sur $[0, 1]$. Il s'est avéré que la vitesse de convergence de la largeur des arbres de Cayley vers le maximum de l'excursion brownienne était égale (à des miettes près) à la vitesse de convergence du maximum du processus empirique vers le maximum du pont brownien. Au passage, on donne une preuve élémentaire du fait que le déplacement total dans une table de hachage uniforme (ou dans un problème de parking usuel) convergeait, normalisé par $n^{3/2}$, vers l'aire sous l'excursion brownienne; on retrouve par là, de manière élémentaire, le résultat de Flajolet & al. [60].

3 Recherche du maximum d'une marche aléatoire

Petit résumé de [14], en commun avec P. Chassaing et M. Yor.

On considère une marche aléatoire $(S_n(k))_{k \in \llbracket 0, n \rrbracket}$ de longueur n , telle que pour $i \geq 1$

$$S_n(i) = X_1 + \cdots + X_i$$

où les $(X_i)_i$ sont i.i.d. et valent ± 1 avec probabilité $1/2$.

On suppose que la marche nous est cachée ; on cherche à trouver le maximum de la marche en s'enquérant de la valeur de $S_n(k)$ pour des valeurs de k de notre choix. Chaque question coûte 1. La question est la suivante :

Quelle stratégie permet de trouver le maximum en posant le moins de questions ?

Odlyzko [92] a trouvé un algorithme qui, en moyenne, atteint la borne inférieure du coût moyen de tous les algorithmes pour le premier terme du développement asymptotique.

Dans ce travail, nous calculons la limite en loi de cet algorithme et exprimons son coût grâce à une fonctionnelle du brownien. D'autre part, on montre que l'algorithme proposé par Odlyzko est également asymptotiquement optimal pour l'ordre stochastique : autrement dit, la fonction de répartition de la limite de l'algorithme d'Odlyzko est une borne supérieure, point par point, pour la limite supérieure des fonctions de répartition du coût de toute suite d'algorithmes.

C'est la première fois, à part cas triviaux, où il est montré qu'un algorithme possède une telle propriété.

— Références —

- [15] A. Akhavi, (2002). *Threshold phenomena in random lattices and efficient reduction algorithms*, Theoretical Computer Science, 257, 359-385.
- [16] A. Akhavi, (1999). *Analyse comparative d'algorithmes de réduction sur les réseaux aléatoires*, Thèse Université de Caen 1999.
- [17] D. Aldous, (1988). *Hashing with linear probing under nonuniform probabilities*. Prob. Eng. Inf. Sci. **2**, 1-14.
- [18] D. Aldous, (1991). *The continuum random tree. I*, Ann. Probabb. 19, 1–28.
- [19] D. Aldous, (1991). *The continuum random tree. II : An overview*, Stochastic analysis, Proc. Symp., Durham/UK 1990, Lond. Math. Soc. Lect. Note Ser. 167, 23-70 .
- [20] D. Aldous, (1993). *The continuum random tree. III*, Ann. Probab. 21, No.1, 248-289 .
- [21] D. Aldous, (1993). *Tree-Based Models for Random Distribution of Mass*, J. Statist. Phys. 73 625-641.
- [22] D. Aldous, (1998). *Brownian excursion conditioned on its local time*, Elect. Comm. in Probab., **3** , 79-90.
- [23] D. Aldous, G. Miermont & J. Pitman, (2004). *The exploration process of inhomogeneous continuum random trees, and an extension of Jeulin's local time identity*, Probab. Theory Related fields, **129** no. 2, 182–218.
- [24] J. Ambjørn, B. Durhuus & J. Fröhlich, (1985). *Diseases of triangulated random surfaces*, Nucl. Phys. B, 2571-433.
- [25] J. Ambjørn, B. Durhuus & T. Jonsson, (1997). *Quantum Geometry. A Statistical Field Theory Approach*, Cambridge.
- [26] J. Ambjørn & Y. Watabiki, (1995). *Scaling in quantum gravity*, Nucl. Phys. B 445, no. 1, 129-142.
- [27] O. Angel & O. Schramm, (2003). *Uniform infinite planar triangulations*, arXiv : math.PR/0207153.
- [28] R. Arratia, A. D. Barbour & S. Tavaré, (2003). *Logarithmic combinatorial structures : a probabilistic approach*, European Mathematical Society (EMS), Zurich, Monographs in Mathematics.
- [29] K.B. Athreya & P.E. Ney, (1972). *Branching processes*, Springer-Verlag.
- [30] J. Bennes, G. Kersting, (2000). *A random walk approach to Galton-Watson trees.*, J. Theor. Probab. 13, No.3, 777-803.
- [31] J. Bertoin, (2000). *A fragmentation process connected to Brownian motion*, Probab. Theory Related Fields, **117**,2, 289–301.
- [32] J. Bertoin, (2003). *The asymptotic behavior of fragmentation processes*, J. Europ. Math. Soc., **5** , 4, 395-416.
- [33] J. Bertoin & G. Miermont (2004). *Asymptotics in Knuth's parking problem for caravans*, preprint, <http://www.dma.ens.fr/~miermont/parking.pdf>.
- [34] J. Bertoin & A. Rouault (2004). *Discretization methods for homogeneous fragmentations.*, <http://front.math.ucdavis.edu/math.PR/0409545>

- [35] J.D. Biggins, (1992). *Uniform convergence of martingales in the branching random walk*, Ann. Probab., **20**, 1, 137–151.
- [36] J. Bouttier, P. Di Francesco & E. Guitter, (2003). *Statistics of planar graphs viewed from a vertex : A study via labeled trees*, Nucl.Phys. B675, 631-660.
- [37] P. Chassaing & B. Durhuus, (2003). *Statistical Hausdorff dimension of labelled trees and quadrangulations*, Arkiv : math.PR/0311532, Nov 2003.
- [38] P. Chassaing & S. Janson, (2001). *A Vervaat-like path transformation for the reflected Brownian bridge conditioned on its local time at 0*, . Ann. Probab. 29, no. 4, 1755–1779.
- [39] P. Chassaing & G. Louchard, (1999). *Phase transition for parking blocks, Brownian excursion and coalescence*. Preliminary version.
- [40] P. Chassaing & R. Marchand, (2004). *Merging costs for the additive Marcus-Lushnikov process, and Union-Find algorithms*, <http://xxx.lanl.gov/abs/math.PR/0406094>.
- [41] P. Chassaing & G. Schaeffer, (2002). *Random planar lattices and integrated superBrownian excursion*, Prob. Theory Rel. Fields., Vol. 128, No. 2, 161 - 212.
- [42] B. Chauvin, M. Drmota & J. Jabbour-Hattab, (2001). *The profile of Binary Search Trees*, Ann. Appl. Prob., 11, 1042-1062.
- [43] B. Chauvin & A. Rouault, (2003). *Connecting Yule process, bisection and binary search trees via martingales*, à paraître dans J. Iranian Stat. Soc., (<http://www.math.uvsq.fr/~rouault/rouaultart/bst3.pdf>).
- [44] R. Cori & B. Vauquelin, (1981). *Planar maps are well labeled trees*. Canad. J. Math. 33(5), 1023-1042.
- [45] H. Daudé & B. Vallée, (1994). *An upper bound on the average number of iterations of the LLL algorithm*. Theoretical Computer Science 123, 95-115.
- [46] L. Devroye, (1998). *Branching processes and their applications in the analysis of tree structures and tree algorithms*. Probabilistic methods for algorithmic discrete mathematics, eds. M. Habib et al., Algorithms Combin. 16, Springer, Berlin, 249-314.
- [47] L. Devroye, (1985). *The expected length of the longest probe sequence for bucket searching when the distribution is not uniform*, J. Algorithms 6, 1-9.
- [48] L. Devroye, J.A. Fill & R. Neininger, (2000). *Perfect Simulation from the Quicksort Limit Distribution*, Elec. Com. Probab., 5, 95-99.
- [49] M. Drmota & B. Gittenberger, (1997). *On the profile of random trees*, Random Structures Algorithms 10, no. 4, 421–451.
- [50] T. Duquesne, (2003). *A limit theorem for the contour process of conditioned Galton-Watson trees*, Ann. Probab. 31, No.2, 996-1027.
- [51] T. Duquesne & J.F. Le Gall, (2002). *Random trees, Lévy processes and spatial branching processes*. Astérisque 281. Soc. Math. de France, Paris.
- [52] T. Duquesne & J.F. Le Gall, (2004). *Probabilistic and fractal aspects of Lévy trees*. Prépublication.
- [53] R. Durrett, H. Kesten & E. Waymire, (1991). *On weighted heights of random trees*. J. Theor. Probab. 4, No.1, 223-237.
- [54] A. Dvoretzky & Th. Motzkin, (1947). *A problem of arrangements*, Duke Math. J. 14, 305–313.

- [55] S.N. Evans, J. Pitman & A. Winter, (2004). *Rayleigh processes, real trees, and root growth with re-grafting*, <http://oz.Berkeley.EDU/users/evans/publications.html>
- [56] W. Feller, (1971). *An introduction to probability theory and its applications. Vol II. 2nd ed.*, Wiley.
- [57] J.A. Fill & S. Janson, (2001). *Approximating the Limiting Quicksort Distribution*, Special Issue of Analysis on Algorithms, RSA, 19, 376-406.
- [58] J.A. Fill & S. Janson, (2002). *Quicksort Asymptotics*, Special Issue of Analysis on Algorithms, RSA, 44, 4-28.
- [59] P. Flajolet & A. Odlyzko, (1982). *The average height of binary trees and other simple trees*, J. Comput. System Sci. **25**, no. 2, 171–213.
- [60] P. Flajolet, P. Poblete & A. Viola, (1998). *On the Analysis of Linear Probing Hashing*. Algorithmica Vol. **22** (4) : 490-515.
- [61] P. Flajolet & R. Sedgewick, (1996). *An introduction to the analysis of algorithms*, Amsterdam : Addison-Wesley, xv.
- [62] J. Geiger & G. Kersting, (1999). *The Galton-Watson tree conditioned on its height*. [CA] Grigelionis, B. (ed.) et al., Probability theory and mathematical statistics. Proceedings of the 7th international Vilnius conference, Vilnius, Lithuania, August, 12-18,. Vilnius : TEV. 277-286.
- [63] B. Gittenberger, (2003). *A note on “State spaces of the snake and its tour – Convergence of the discrete snake” by J.-F. Marckert and A. Mokkadem*, J. Theo. Probab. 16, No. 4, 1063-1067.
- [64] G.H. Gonnet, (1980). *Open-addressing hashing with unequal-probability keys*. J. Comput. Syst. Sci. 21, 354-367.
- [65] T.E. Harris, (1952). *First passage and recurrence distributions*, Trans. Amer. Math. Soc. 73, 471-486.
- [66] H.K. Hwang, (1995). *Asymptotic expansions for the Stirling numbers of the first kind*, J. Combin. Theory Ser. A, **71** , 2, 343–351.
- [67] J. Jabbour-Hattab, (2001). *Martingales and Large Deviations for Binary Search Trees*, Random Structure and Algorithms, 19, 112-127.
- [68] S. Janson, (2001). *Asymptotic distribution for the cost of linear probing hashing*, Random Struct. Alg. 19, no. 3-4, 438-471.
- [69] S. Janson, (2003). *Individual displacements for linear probing hashing with different insertion policies*, Tech. Report 35, Uppsala.
- [70] W.D. Kaigh, (1976). *An invariance principle for random walk conditioned by a late return to zero*, Ann. Probab. 4, 115-121.
- [71] O. Kallenberg, (2002). *Foundations of Modern Probability*. 2nd ed., Probability and Its Applications. Springer, New York, NY.
- [72] D.G. Kendall, (1951). *Some problems in the theory of queues*, J. Royal Stat. Soc., B 13, 151-185.
- [73] D.P. Kennedy, (1976). *The distribution of the maximum Brownian excursion*, J. Appl. Probab. **13**, 371-376.
- [74] H. Kesten, (1994). *A limit theorem for weighted branching process trees*, in The Dynkin Festschrift. Markov processes and their applications, Freidlin ed. Birkhäuser, 153-166.
- [75] J. Komlos, P. Major & G. Tusnady, (1976). *An approximation of partial sums of independent RV's and the sample DF. II.*, Z. Wahrs. verw. **34**, 33-58.

- [76] D. E. Knuth, (1997). *The art of computer programming. Vol. 1 : Fundamental algorithms. 3rd ed.*, Addison-Wesley
- [77] D. E. Knuth, (1997). *The Art of Computer Programming. Vol. 3 : Sorting and Searching.* 2nd ed., Addison-Wesley.
- [78] T. Jeulin, (1980). *Semi-martingales et grossissement d'une filtration*, Lecture Notes in Math., 833, Springer-Verlag.
- [79] J.F. Le Gall & Y. Le Jan, (1995). *Arbres aléatoires et processus de Lévy. (Random trees and Lévy processes)*, C. R. Acad. Sci., Paris, Sér. I 321, No.9, 1241-1244.
- [80] J.F. Le Gall & Y. Le Jan, (1998). *Branching processes in Lévy processes : The exploration process.* Ann. Probab. 26, No.1, 213-252.
- [81] J.F. Le Gall, (1999). *Spatial branching processes, random snakes and partial differential equations*, Lectures in Mathematics, Birkhäuser.
- [82] A.K. Lenstra, H.W. Lenstra & L. Lovász (1982). *Factoring polynomials with rational coefficients*, Math. Ann. 261, 513–534.
- [83] V. Limic, (2000) *On the behavior of LIFO preemptive resume queues in heavy traffic.* Electron. Commun. Probab. 5, 13-27, .
- [84] V. Limic, (2001) *A LIFO queue in heavy traffic*, Ann. Appl. Prob. 11, No.2, 301-331.
- [85] V.A. Liskovets, (1981). *A census of nonisomorphic planar maps, Algebraic methods in graph theory, Vol. I, II*, Colloq. Math. Soc. János Bolyai, 25, 479–494, North-Holland, Amsterdam,
- [86] V. A. Liskovets, (1981). *Enumeration of nonisomorphic planar maps*, J. Graph Theory, V5, 1, 115–117
- [87] R. Lyons with Y. Peres, (2005 ??). *Probability on Trees and Networks*, <http://mypage.iu.edu/~rd-lyons/prbtree/prbtree.html>
- [88] R. Lyons, R. Pemantle, & Y. Peres, (1995). *Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes*, Ann. Probab. 23, No.3, 1125-1138.
- [89] H. Mahmoud, (1992). *Evolution of Random Search Trees*, John Wiley.
- [90] R. Neininger & L. Rüschendorf, (2004). *A general limit theorem for recursive algorithms and combinatorial structures*, Annals of App. Probab., 14, 1, 378–418.
- [91] J. Neveu, (1986). *Arbres et processus de Galton-Watson*, Annales de l'Inst. H. Poincaré, Vol. 22, 2.
- [92] A.M. Odlyzko, (1995). *Search for the maximum of a random walk*, Random Structures and Algorithms, 6, 275-295.
- [93] A.M. Odlyzko & H.S. Wilf, (1987). *Bandwidths and profiles of trees*, J. Comb. Theory, Ser. B 42, 348-370.
- [94] V.V. Petrov, (1975). *Sums of independant random variables*, Springer .
- [95] G.CH. Pflug & H.W. Kessler, (1987). *Linear probing with a nonuniform address distribution*, Journal of ACM, 34 (2), 397-410.
- [96] J. Pitman, (1999). *The SDE solved by local times of a Brownian excursion or bridge derived from the height profile of a random tree or forest*, Ann. Probab; 27 , 261–283.
- [97] J. Pitman, (2002). *Combinatorial Stochastic Processes. Lecture Notes for St. Flour Summer School.* Available at : <http://stat-www.berkeley.edu/users/pitman/bibliog.html>

- [98] B. Pittel, (1987). *Linear probing : the probable largest search time grows logarithmically with the number of records*. J. Algorithms, Vol. **8** (2) : 236–249.
- [99] P.V Poblete & A. Viola, (2001). *The effect of deletions on different insertion disciplines for hash tables*, Szwarcfiter, Jayme (ed.), Proceedings of the Brazilian symposium on graphs, algo. and comb., Fortaleza, Ceará, Brazil, March 17-19. Extended abstracts. Elsevier, Electron. Notes Discrete Math. 7, no pag., electronic only (2001).
- [100] B. Reed, (2003). *The height of a random binary search tree*, Journal of the ACM, **50** (3), 306-332.
- [101] J.M. Robson, (2002). *Constant bounds on the moments of the height of binary search trees*, Theor. Computer Sci, 276, 435–444.
- [102] U. Rösler, (1991). *A limit theorem for “Quicksort”*, RAIRO, Inform. Théor. Appl., 25, 1,85-100.
- [103] U. Rösler, (2001). *On the analysis of stochastic divide and conquer algorithms*, Algorithmica, 29, 1-2, 238–261.
- [104] G. Schaeffer, (1998). *Conjugaison d’arbres et cartes combinatoires aléatoires*, PhD Thesis, Université Bordeaux I.
- [105] L. Serlet, (1997). *A large deviation principle for the Brownian snake*, Stochastic Processes Appl. 67, No.1, 101-115.
- [106] J. Spencer, (1997). *Enumerating graphs and Brownian motion*. Commun. Pure Appl. Math. Vol. **50** (3), 291-294.
- [107] L. Takács, (1993). *Limit distributions for queues and random rooted trees*, J. Appl. Math. Stoch. Ana., **6** , No.3, 189-216.
- [108] W.T. Tutte, (1963). *A census of planar maps*, Canad. J. Math., 15, p. 249-271.
- [109] K. Uchiyama, (1982). *Spatial growth of a branching process of particles living in R^d* , Ann. Probab., **10**, 4, 896–918.
- [110] B. Vallée, (1989). *Un problème central en géométrie algorithmique des nombres : la réduction des réseaux. Autour de l’algorithme LLL*, Informatique Théorique et Applications 3 , 345–376.
- [111] Y. Watabiki, (1995). *Construction of non-critical string field theory by transfer matrix formalism in dynamical triangulation*, Nucl. Phys. B 441, No.1-2, 119-163.