



HAL
open science

Structures arborescentes aléatoires

Etienne Bellin

► **To cite this version:**

Etienne Bellin. Structures arborescentes aléatoires. Probabilités [math.PR]. Institut Polytechnique de Paris, 2023. Français. NNT : 2023IPPAX037 . tel-04457233

HAL Id: tel-04457233

<https://theses.hal.science/tel-04457233>

Submitted on 14 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2023IPPAX037

Thèse de doctorat



Structures arborescentes aléatoires

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'École polytechnique

École doctorale n°574 École doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 7 juillet 2023, par

ETIENNE BELLIN

Composition du Jury :

Philippe Biane Directeur de recherche, Université Paris-Est (IGM)	Président
Nicolas Broutin Professeur, Sorbonne Université (LPSM)	Rapporteur
Luc Devroye Professeur, McGill University (SoCS)	Rapporteur
Bénédicte Haas Professeur, Université Paris 13 (LAGA)	Examineur
Igor Kortchemski Chargé de recherche, École polytechnique (CMAP)	Directeur de thèse
Lucas Gerin Professeur assistant, École polytechnique (CMAP)	Examineur

STRUCTURES ARBORESCENTES ALÉATOIRES

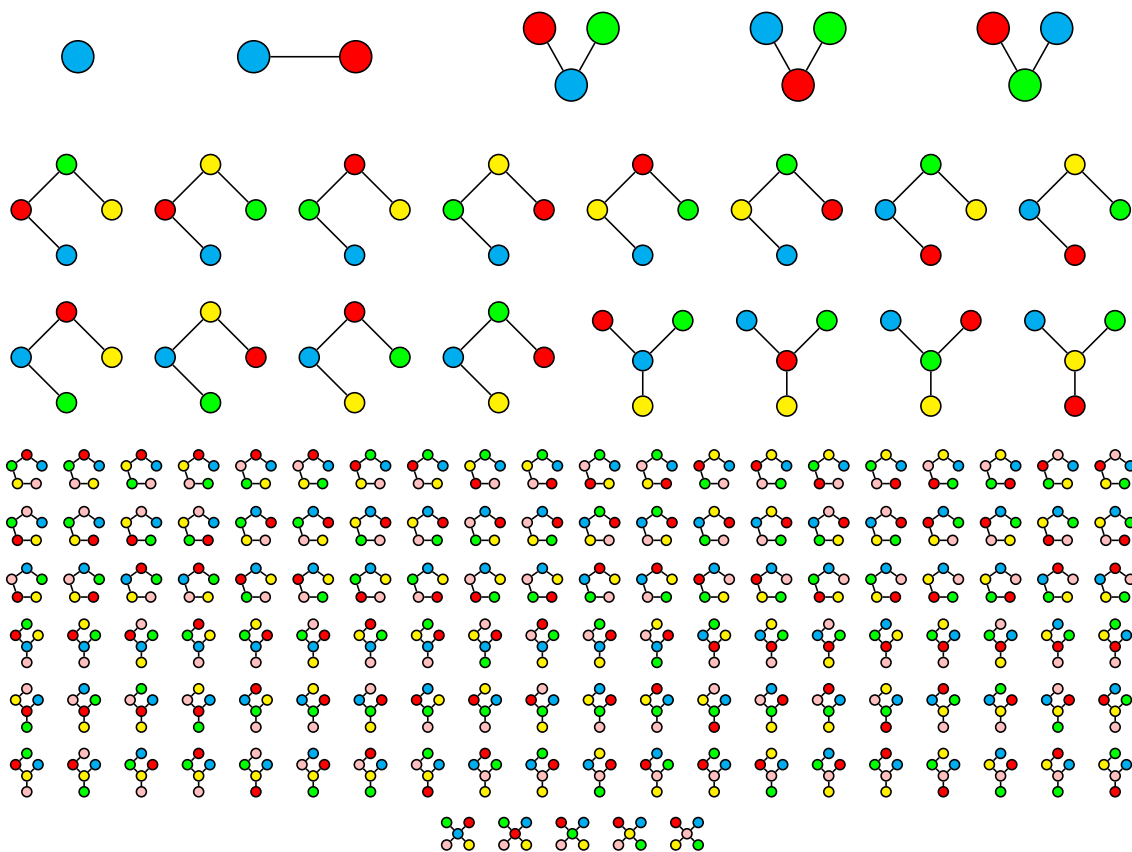


FIGURE 1 – Tous les arbres de Cayley à 1,2,3,4 et 5 sommets.



REMERCIEMENTS

MES PREMIERS REMERCIEMENTS iront bien évidemment à Igor qui a encadré ma thèse et a su m'aiguiller dans le monde de la recherche. Il fut de très bons conseils dans le domaine mathématiques mais, aussi, sur bien d'autres sujets. Je prends conscience de la chance que j'ai eu de travailler avec lui.

Je voudrais aussi remercier l'École Polytechnique et le Centre de Mathématiques Appliquées grâce à qui j'ai pu assister à différentes conférences et élargir mes connaissances mathématiques et humaines.

Une pensée à tous les doctorants du CMAP que j'ai rencontrés et avec qui j'ai discuté. Bon courage à tous ceux qui soutiendront après moi. C'est une chance inouïe de pouvoir faire une thèse ici.

Merci à mes amis qui m'ont épaulé et à qui je peux me confier sans crainte. Un mot spécial à un certain logicien que je connais bien : nous n'avons jamais publié d'articles ensemble (même si j'attends toujours de pouvoir combiner logique et probabilités) mais tu restes mon plus grand collaborateur.

Je tiens à remercier ma famille qui m'a soutenu durant cette longue entreprise. Ils me comprennent mieux que quiconque et sont la raison pour laquelle j'en suis ici aujourd'hui. Je tiens à les rendre fiers par ce travail car je sais que c'est à leurs yeux qu'il aura le plus de valeur.

Enfin, je voudrais rendre hommage à la femme qui m'a rejoint en cours de route et avec qui je partage ma vie depuis. Elle est celle qui me connaît le mieux et celle à qui j'aimerais dédier ce travail.

À Mathilde,



RÉSUMÉ

Cette thèse se concentre sur certaines structures arborescentes aléatoires et leurs applications. Quatre modèles sont étudiés. Le premier est celui des factorisations minimales du cycle de taille n . Une telle factorisation est la donnée de $n - 1$ transpositions dont la composition égale la permutation cyclique $(1 \cdots n)$. En exploitant une bijection avec les arbres étiquetés et en explicitant certaines fonctions génératrices sur ces arbres, nous déduisons la loi jointe du nombre de fois que 1 et 2 apparaissent dans une factorisation minimale de taille n choisie uniformément au hasard. Le deuxième problème est celui des tuples de stationnement (*parking functions*). Nous étudions la distance entre les k premières coordonnées d'un n -uplet de stationnement choisi uniformément au hasard et k variables i.i.d uniformes. Nous montrons que la distance en variations totales tend vers 0 lorsque $k = o(\sqrt{n})$ et que la distance de Kolmogorov tend vers 0 lorsque $k = o(n)$. Nous donnons aussi des estimations sur la vitesse de convergence. Cette étude est possible, encore une fois, grâce à une bijection avec les arbres étiquetés qui permet de se ramener à l'étude d'arbres aléatoires puis de marches aléatoires conditionnées. Le troisième problème consiste à étudier le nombre d'indépendance d'arbres aléatoires. Le nombre d'indépendance d'un graphe est le plus grand cardinal d'un sous-ensemble de sommets du graphe tel qu'aucune paire de sommets de cet ensemble ne sont reliés par une arête. Nous obtenons des théorèmes limites pour les arbres simplement générés par des méthodes robustes de convergence locale, et généralisons des résultats précédemment obtenus de manière calculatoire. Enfin le dernier problème est celui des factorisations monotones qui sont des factorisations minimales satisfaisant une certaine condition supplémentaire de monotonie. Nous décrivons une bijection entre ces objets et les arbres plans. Pour tout $k < n$ nous représentons les k premières transpositions d'une factorisation monotone de taille n comme une lamination du disque, i.e comme une union de cordes disjointes du cercle unité. Nous étudions ensuite la convergence du processus $(0 \leq k < n)$ de lamination en exploitant la bijection avec les arbres plans.



Table des matières

Introduction	1
I Structures Arborescentes et Applications	13
A Arborescences finies	14
B Arborescences ordonnées	23
C Arborescences aléatoires	33
D Convergence de grands arbres	39
II Factorisations Minimales du Cycle	55
A Introduction	56
B Bijection between minimal factorizations and Cayley trees	57
C First generating function on Cayley trees	60
D A second generating function on Cayley trees and Proof of Theorem II.A.1	64
III Tuples de Stationnement	69
A Introduction	70
B Bijection between parking functions and Cayley trees	72
C Convergence for the total variation distance	74
D Convergence for the Kolmogorov distance	81
E Sum and maximum of the first parking places	84
IV Plus Grand Ensemble d'Indépendants	87
A Introduction	88
B Simply generated trees	90
C Main results	91
D Limit theorems for uniformly pointed simply generated trees	92
E Properties of the tricolouration	95
F Tricolouration of the infinite limiting trees	96
G Proof of Theorem IV.C.1	98
V Factorisations Primitives du Cycle	103
A Introduction	104
B Bijections between plane trees, decreasing and increasing factorisations	107
C Asymptotic behaviour of random uniform decreasing factorisations	116





INTRODUCTION

COMMENÇONS par analyser brièvement le titre de cette thèse afin d'en apercevoir le contenu. Les *structures arborescentes* font référence à tout modèle pouvant se représenter sous forme d'un arbre ou d'un objet semblable. Les *arbres* sont des objets mathématiques très simples que l'on peut rencontrer dans de nombreux contextes même en dehors du domaine mathématique. Les arbres généalogiques, les arbres de décisions, les organigrammes, les arbres de probabilités, et bien évidemment, les végétaux éponymes sont de fameux exemples de telles structures. Les arbres ont de nombreuses applications scientifiques notamment en biologie dans la modélisation de l'évolution de populations. En informatique, ils constituent une structure de données importante. De manière générale, les structures arborescentes permettent facilement de représenter toute forme de hiérarchie naturelle entre des éléments. Notre but durant cette thèse sera de mener l'étude de structures arborescentes dotées d'un certain aléa et de l'appliquer à divers problèmes. Pour être plus précis, les problèmes que nous aborderons seront celui des factorisations minimales et monotones du cycle, celui des tuples de stationnement et celui du plus grand ensemble d'indépendants dans un graphe. Ci-dessous nous donnons une approche « fil rouge » qui nous permettra d'aborder leur résolution.

Un fil rouge. Une philosophie qui nous guidera durant toute cette thèse est la suivante. Supposons que l'on doive étudier le comportement asymptotique d'objets discrets. Pour cela nous les codons, c'est-à-dire, en termes plus mathématiques, nous trouvons une bijection, avec d'autres objets discrets dont le comportement est plus accessible. Plus précisément, nous supposons que ces objets alternatifs convergent vers un certain objet limite. Nous souhaitons ensuite appliquer notre codage inverse, ou bijection réciproque, à cet objet en espérant trouver la limite correspondant aux objets de départ. Voir Figure 3 ci-dessous. Les arbres sont au coeur de cette thèse et constitueront tantôt les objets discrets de départ, c'est-à-dire le sujet même de notre étude, tantôt les objets intermédiaires, c'est-à-dire l'outil par lequel mener notre analyse. Parfois, ils constitueront aussi les objets limites. Autrement dit les arbres pourront apparaître au quatre coins de la Figure 3. Les difficultés dans l'application d'une telle stratégie peuvent venir de trois endroits différents.

— (?¹) La première est la description de l'objet limite.

— (?²) La deuxième est l'application de la bijection réciproque à l'objet limite qui, bien souvent, n'est pas de

même nature que les éléments discrets dont il vient. Le codage doit alors s'adapter ce qui n'est pas toujours évident. Cela n'est cependant pas la difficulté majeure en général.

— (?³) La troisième est de montrer que la convergence a bien lieu.

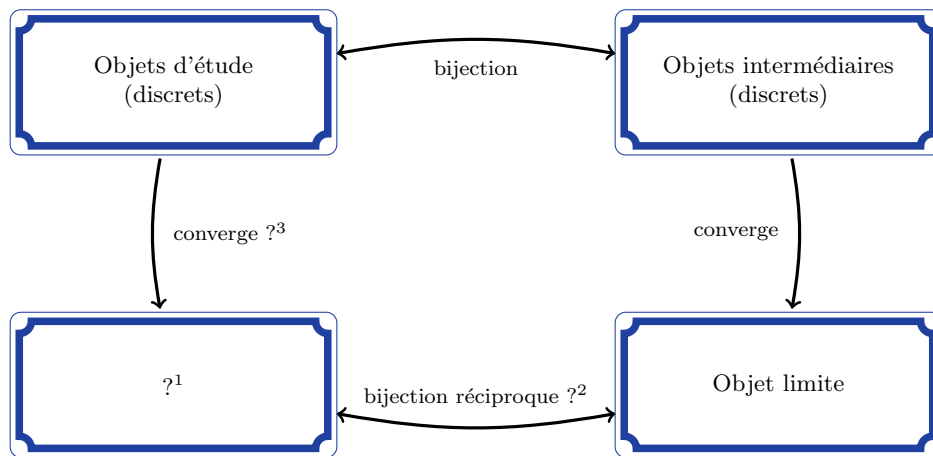


FIGURE 2 – Illustration de la philosophie fil rouge.

Indications de lecture.

Théorème : Résultats connus



Dans le premier chapitre, les résultats déjà connus dans la littérature seront annoncés par le symbole d'une main en haut à droite.

Théorème : Résultats nouveaux



Dans le premier chapitre, les résultats nouveaux démontrés dans cette thèse seront annoncés par le symbole de l'étoile.

Le premier chapitre constitue un résumé en français des travaux effectués durant la thèse. Il ne contient pas les preuves des résultats nouveaux. Toutes les preuves et les détails sont situés dans les chapitres suivants qui constituent une compilation des articles rédigés pendant la thèse.

- Le chapitre 2 est basé sur l'article *Degrees in random uniform minimal factorizations* [14].
- Le chapitre 3 est basé sur *Asymptotic behaviour of the first positions of uniform parking functions* [17].
- Le chapitre 4 est basé sur *On the independence number of random trees via tricolourations* [15].
- Le chapitre 5 est basé sur *Random monotone factorisations of the cycle* [16].

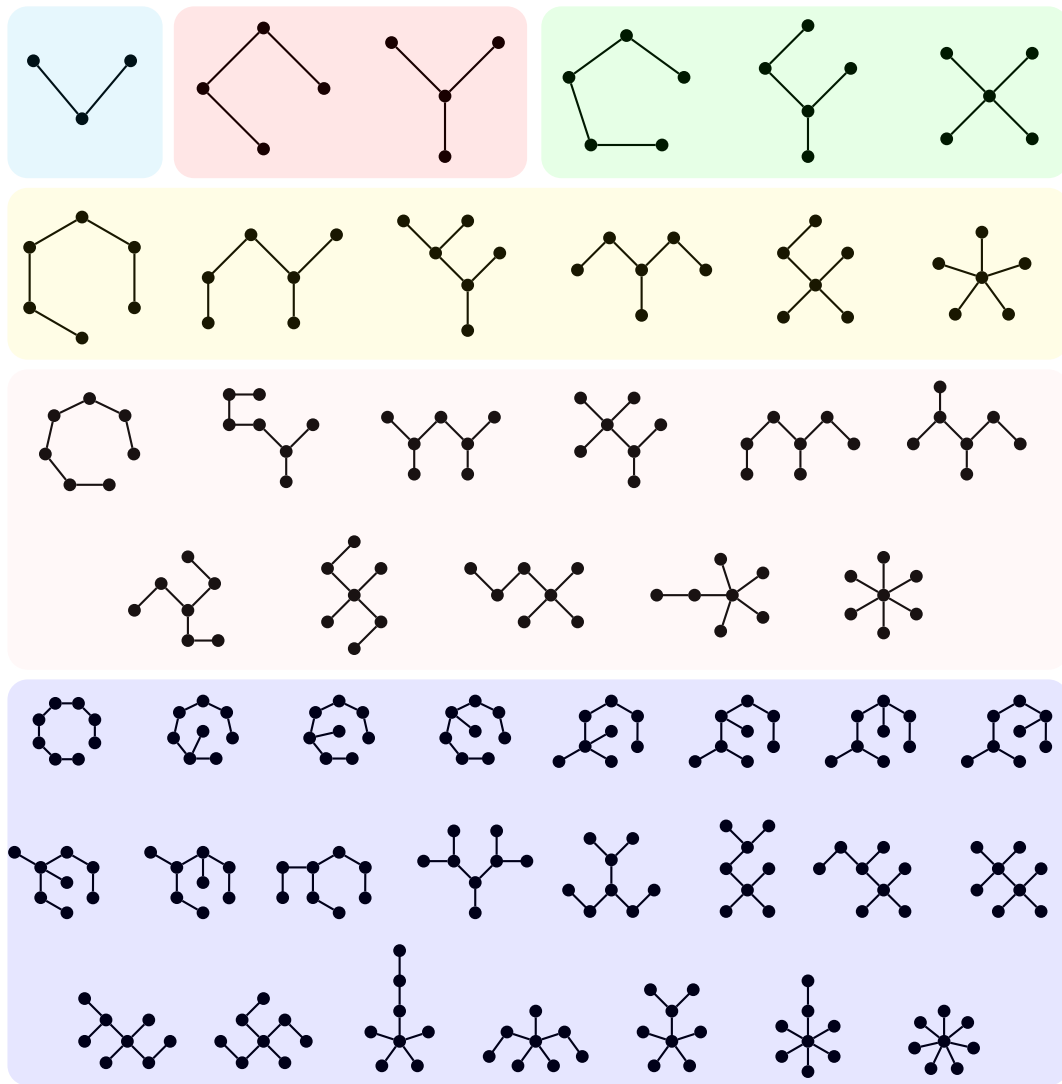


FIGURE 3 – Tous les arbres à 3,4,5,6,7 et 8 sommets à isomorphisme près.



CHAPITRE

I

STRUCTURES ARBORESCENTES ET
APPLICATIONS

Contenu du Chapitre

A Arborescences finies	14
A.i Prolégomènes	14
A.ii Application aux factorisations minimales du cycle	18
B Arborescences ordonnées	23
B.i Prolégomènes	23
B.ii Parcourir un arbre plan fini et enraciné	25
B.iii Application aux tuples de Stationnement	27
B.iv Application aux factorisations monotones du cycle	30
C Arborescences aléatoires	33
C.i Arbres de Bienaymé	33
C.ii Application aux tuples de Stationnement (suite)	35
C.iii Arbres simplement générés	37
D Convergence de grands arbres	39
D.i Limites locales	39
D.ii Application au problème du plus grand ensemble d'indépendants	42
D.iii Limites d'échelles	48
D.iv Application aux factorisation monotones du cycle (suite)	49



A Arborescences finies

A.i Prolégomènes

Définition I.A.1 : Graphe et arbre

Un *graphe* est un couple $G = (S, A)$ où S est un ensemble quelconque et A est un ensemble de paires $\{u, v\}$ avec $u, v \in S$ et $u \neq v$. Les éléments de S sont appelés les *sommets* de G et les éléments de A sont appelés les *arêtes* de G .

Un *arbre* est un graphe $T = (S, A)$ connexe et acyclique, c'est-à-dire que, pour tout $u, v \in S : u \neq v$ si, et seulement si, il existe $n \geq 2$ et $u = u_1, \dots, u_n = v \in S$ tels que $\{u_i, u_{i+1}\} \in A$ pour tout $i \in \llbracket 1, n-1 \rrbracket$.

Remarque

Nous considérons ici uniquement des graphes simples (i.e sans arêtes multiples) et non-orientés. En réalité nous ne parlerons que très peu de graphes et resterons concentrés principalement sur les arbres. Il y a deux raisons à cela. La première est que les arbres constituent déjà une classe assez grande d'objets pouvant modéliser de nombreux phénomènes, comme mentionné dans l'introduction. La deuxième est que les graphes sont généralement plus complexes que les arbres à cause de la présence de cycles. Se contenter des arbres permet souvent de simplifier l'étude. Cela se verra notamment dans la Section D.ii où nous traiterons le problème du plus grand ensemble d'indépendants d'un graphe : le problème se simplifie pour les arbres grâce à un certain procédé de tricoloriage des sommets propre aux arbres.

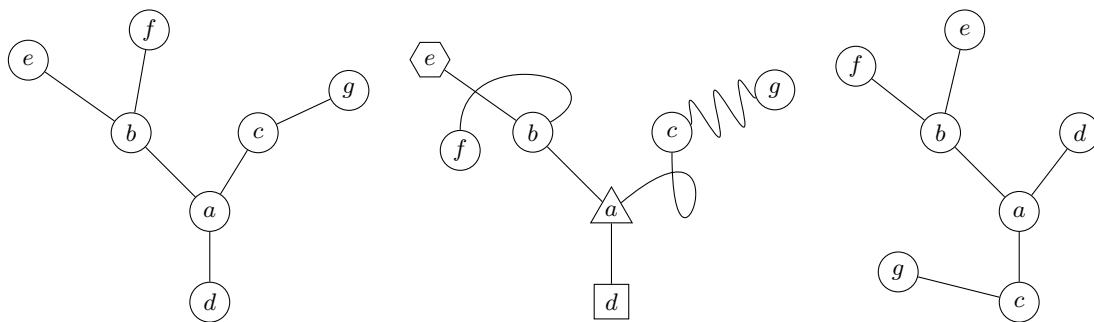


FIGURE A.1 – Trois manières équivalentes de représenter le même arbre $T = (S, A)$ où $S = \{a, b, c, d, e, f, g\}$ et $A = \{\{a, b\}, \{b, e\}, \{b, f\}, \{a, d\}, \{a, c\}, \{c, g\}\}$. La forme et la disposition des arêtes et des sommets n'importent pas. Pour les arbres ordonnés, que nous verrons à la Définition I.B.1, la disposition des sommets sera importante.

Le vocabulaire associé aux graphes est très riche et plutôt que de définir chaque terme séparément dans une liste interminable nous donnons ici les éléments les plus importants à la volée. Si G est un graphe et u un sommet de G , ce que l'on appelle le *degré* de u , noté $\deg(u)$, est le nombre de sommets de G *adjacents* à u , c'est-à-dire, le nombre de sommets v de G tels que $\{u, v\}$ est une arête de G . Le degré de u correspond donc au nombre de *voisins* de u . Un sommet est dit être une *feuille* s'il est de degré 1 et il est dit être *interne* sinon. À noter que, pour un arbre avec au moins deux sommets, chaque sommet interne a un degré au moins égal à 2 (en raison de la connexité de l'arbre). La *distance* entre deux sommets u et v d'un même graphe correspond au

nombre d'arêtes d'un plus court chemin entre u et v (dans un arbre, il n'y a qu'un seul plus court chemin).

Définition I.A.2 : Graphe enraciné

Un *graphe enraciné* est un couple (G, r) tel que G est un graphe et r est un sommet de G . Le sommet r est alors appelé la *racine* de G .

À partir du moment où un arbre est enraciné, il est possible de considérer une relation d'affiliation entre les sommets : un sommet u est dit être le *fil*s d'un sommet v s'il lui est adjacent et si sa distance par rapport à la racine est plus grande (de +1) que celle entre v et la racine. Le sommet v est alors naturellement appelé le *sommet père* ou *parent* de u . La notion de *descendant* et d'*ancêtre* se comprend alors très bien. Dans tous nos dessins, les racines des arbres enracinés seront mises en avant par un double cercle (voir Figure A.2).

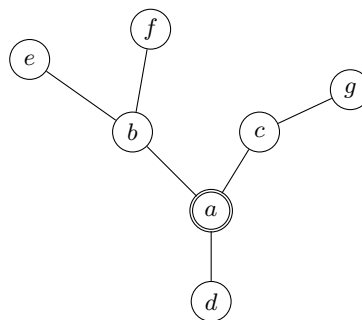


FIGURE A.2 – Représentation de l'arbre enraciné (T, a) où T est l'arbre de la Figure A.1. Ici, la racine est le sommet a qui est mis en avant par un double cercle. Les sommets b , c et d sont les enfants de a . Le sommet b est le parent de e et f . Le sommet g est un descendant de a .

Dans les définitions d'arbre et d'arbre enraciné, rien n'est spécifié pour S , notamment rien n'est supposé sur le cardinal de S . Le cas où S est fini est intéressant car mène à des considérations de dénombrement. D'ailleurs, l'un des premiers résultats, dû à Cayley [32], connus sur les arbres est le nombre d'arbres (enracinés ou non) à n sommets. Plus précisément

Définition I.A.3 : Arbre de Cayley

Un *arbre de Cayley* est un arbre $T = (S, A)$ où $S = \llbracket 1, n \rrbracket$ pour un certain entier $n \in \mathbb{N}$. On note \mathcal{C}_n l'ensemble des arbres de Cayley à n sommets.

Théorème I.A.4 : Dénombrement des arbres de Cayley



Soit $n \geq 2$, alors le nombre d'arbres de Cayley à n sommets vaut n^{n-2} et le nombre d'arbres de Cayley enracinés à n sommets vaut n^{n-1} .

Il existe de nombreuses preuves du Théorème I.A.4 (voir par exemple [4]). L'une d'entre elles utilise ce que l'on appelle le *codage de Prüfer* des arbres enracinés [89]. Ce codage nous sera utile dans la preuve de l'un



de nos résultats (preuve du Lemme III.C.1) c'est pourquoi nous prenons le temps de le décrire ci-dessous. À partir d'un arbre de Cayley T à n sommets, on construit un $(n-2)$ -uplet $(x_1, \dots, x_{n-2}) \in \llbracket 1, n \rrbracket^{n-2}$ en retirant successivement les feuilles de l'arbre T de plus petite valeur et en notant la valeur du sommet parent. Plus précisément on construit la suite d'arbres enracinés $T = T_0, T_1, \dots, T_{n-2}$ de la manière suivante. À l'étape $i \in \llbracket 1, n-2 \rrbracket$, on obtient T_i en retirant à T_{i-1} sa plus petite feuille u , ainsi que l'arête la reliant à son unique voisin v . On pose alors $x_i := v$. Voir Figure A.3. De cette manière on obtient un vecteur (x_1, \dots, x_{n-2}) qui est le codage de Prüfer de l'arbre T . La preuve du Théorème I.A.4 tient dans le fait que ce codage est bijectif. En réalité, dans la preuve du Lemme III.C.1 nous utiliserons une variante du codage de Prüfer où, au lieu de retirer les feuilles de plus petite valeur, nous retirerons les feuilles de plus grande valeur. Cela ne change évidemment rien au côté bijectif du codage.

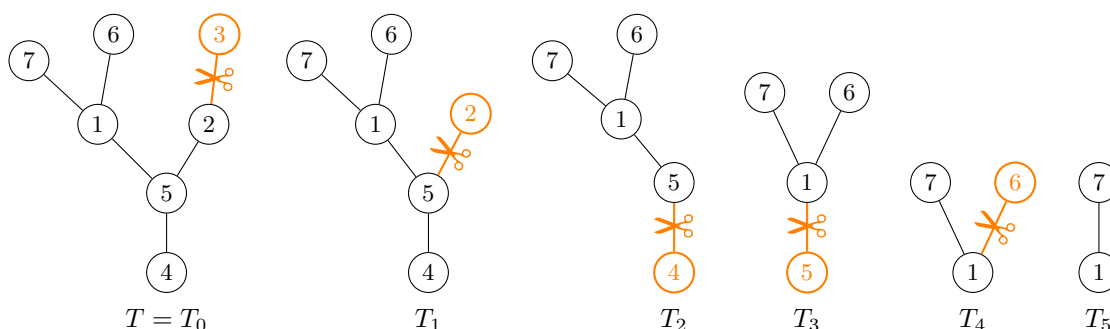


FIGURE A.3 – Illustration du codage de Prüfer. Dans cet exemple, on part d'un arbre de Cayley avec $n = 7$ sommets. Le codage de Prüfer donne le 5-uplet $(2, 5, 5, 1, 1)$. À chaque étape i , on a colorié la feuille de plus petite valeur en orange. L'algorithme s'arrête lorsqu'il ne reste plus que 2 sommets.

Citons un autre résultat bien connu sur les arbres de Cayley qui concerne la fonction génératrice des degrés des sommets. Plus précisément, pour $t \in \mathcal{C}_n$ et $i \in \llbracket 1, n \rrbracket$, notons $\text{deg}_i(t)$ le degré du sommet i dans t . On définit alors la fonction génératrice

$$F_n(x, y) := \sum_{t \in \mathcal{C}_n} x^{\text{deg}_1(t)} y^{\text{deg}_2(t)}.$$

Remarque

La fonction F_n n'est pas réellement la fonction génératrice de $(\text{deg}_1, \text{deg}_2)$ puisqu'il faudrait renormaliser par le cardinal de \mathcal{C}_n , mais pour des raisons de simplicité, nous ignorerons ce terme de renormalisation et continuerons à appeler F_n « fonction génératrice ».

Si l'on pioche un élément T de \mathcal{C}_n uniformément au hasard, alors la fonction génératrice F_n contient toute l'information sur la loi jointe $(\text{deg}_1(T), \text{deg}_2(T))$. Il est bien connu (voir par exemple [18]) que F_n possède une forme explicite très simple.

Théorème I.A.5 : Fonction génératrice des degrés



Pour tout $n \geq 2$, $F_n(x, y) = xy(x + y + n - 2)^{n-2}$.

Remarque

En réalité, le Théorème I.A.5 se généralise en prenant le degré de tous les sommets et non pas juste celui de 1 et 2. Plus précisément on a pour tout $n \geq 2$

$$\sum_{t \in \mathcal{C}_n} \prod_{i=1}^n x_i^{\deg_i(t)} = x_1 \cdots x_n (x_1 + \cdots + x_n)^{n-2}.$$

Notre première contribution sera de généraliser le Théorème I.A.5 en ajoutant un troisième paramètre particulier que nous définissons maintenant. Soit $t \in \mathcal{C}_n$ et $i \in \llbracket 1, n \rrbracket$. Considérons la plus grande suite possible de sommets v_1, \dots, v_ℓ telle que

- Les sommets v_1, \dots, v_ℓ forment un chemin, c'est-à-dire que, pour tout $k \in \llbracket 1, \ell - 1 \rrbracket$, v_k et v_{k+1} sont adjacents.
- Ce chemin part de i , autrement dit, $v_1 = i$.
- Le chemin est croissant dans le sens où $v_{k+1} > v_k$ pour tout $k \in \llbracket 1, \ell - 1 \rrbracket$.
- Enfin, à chaque étape $k \in \llbracket 1, \ell - 1 \rrbracket$, le sommet v_{k+1} est le plus petit possible parmi tous les sommets adjacents à v_k (tout en restant plus grand que v_k afin de respecter le point précédent).

Le chemin ainsi obtenu est noté $L_i(t)$ et sa taille, qui correspond au nombre d'arêtes formant le chemin, est notée $|L_i(t)| := k - 1$ (voir Figure A.4). Le chemin $L_i(t)$ peut être vu comme ce que donnerait un algorithme glouton dont le but serait de maximiser la longueur d'un chemin croissant partant de i . La quantité $|L_1(t)|$ est

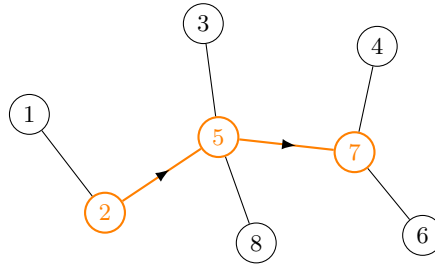


FIGURE A.4 – Exemple d'un arbre $t \in \mathcal{C}_8$. Le chemin $L_2(t) = (2, 5, 7)$ partant de 2 est représenté en orange. Sa taille est $|L_2(t)| = 2$.

intéressante car, comme nous le verrons Section A.ii, elle est reliée aux trajectoires des factorisations minimales. Élargissons la définition de F_n en y incluant le nouveau paramètre associé à $|L_1|$:

$$F_n(x, y, z) := \sum_{t \in \mathcal{C}_n} x^{\deg_1(t)} y^{\deg_2(t)} z^{|L_1(t)|}.$$

On obtient alors le résultat suivant.

Théorème I.A.6 : B. [14]



Pour tout $n \geq 3$,

$$F_n(x, y, z) = xyz \left[x(n-2+x)^{n-3} \left(1 - \frac{yz}{z+y-1} \right) + (x+y+z+n-3)^{n-3} \left(n-2+yz + \frac{xyz}{z+y-1} \right) \right]. \quad (\text{A.1})$$

De manière assez surprenante la formule (A.1) est symétrique en y et en z ce qui implique le corollaire suivant.



Corollaire I.A.7 : B. [14]

Soit T_n un arbre de Cayley aléatoire à n sommets de loi uniforme, alors $(\deg_1(T_n), |L_1(T_n)|, \deg_2(T_n))$ et $(\deg_1(T_n), \deg_2(T_n), |L_1(T_n)|)$ ont la même loi.

Cela répond positivement à une conjecture de Caraceni [30]. Il serait néanmoins intéressant de trouver une manière bijective de démontrer le Corollaire I.A.7 car la preuve du Théorème I.A.6 ne donne pas la raison intuitive de cette symétrie. Comme mentionné plus haut, la raison pour laquelle nous étudions la fonction génératrice trivariée F_n est qu'elle permet d'étudier les propriétés d'une famille d'objets combinatoires, appelés factorisations minimales, qui est en bijection avec les arbres de Cayley. Cette étude est le sujet de la prochaine section.

A.ii Application aux factorisations minimales du cycle

Les arbres de Cayley sont en bijection avec de nombreuses familles d'objets combinatoires. Leur étude permet donc de déduire des propriétés intéressantes sur ces dernières. Un exemple d'une telle famille est donné par les factorisations minimales du cycle dont voici la définition.

Définition I.A.8 : Factorisation minimale

Soit un entier $n \geq 2$. On considère le cycle $(1\ 2\ \dots\ n)$, qui est une permutation de \mathfrak{S}_n . Une *factorisation minimale* de ce cycle est la donnée de $n-1$ transpositions $(\tau_1, \dots, \tau_{n-1})$ telles que

$$\tau_{n-1} \circ \dots \circ \tau_1 = (1\ 2\ \dots\ n).$$

On note \mathcal{F}_n l'ensemble des factorisations minimales du cycle de taille n .

Par exemple,

- Le $(n-1)$ -uplet $((n\ n-1), (n-1\ n-2), \dots, (3\ 2), (2\ 1))$ est toujours une factorisation minimale de $(1\ 2\ \dots\ n)$. Cela montre déjà que \mathcal{F}_n n'est jamais vide.
- Le 9-uplet $((8\ 9), (8\ 10), (7\ 8), (2\ 3), (2\ 4), (2\ 5), (1\ 2), (1\ 6), (1\ 7))$ est une factorisation minimale de $(1\ \dots\ 10)$ qui n'est pas de la forme ci-dessus.

Motivations et bibliographie. Les factorisations minimales sont intéressantes car elles sont en connexion avec d'autres objets combinatoires comme les tuples de stationnement [94, 20] que nous verrons plus tard (cf Section B.iii), ou encore, les partitions non croisées [19]. La manière la plus simple d'apporter un point de vue probabiliste à l'étude d'une famille d'objets combinatoires est de piocher au hasard un des éléments dans cette famille de manière uniforme. Dans ce contexte-ci, des ponts se forment avec d'autres structures aléatoires comme la triangulation brownienne [51] et la fragmentation d'Aldous-Pitman de l'arbre brownien [98].

Les factorisations de \mathcal{F}_n sont qualifiées de « minimales » car il s'avère qu'il faut au moins $n-1$ transpositions pour espérer décomposer le cycle $(1\ 2\ \dots\ n)$. Autrement dit, la composition de k transpositions, avec $k < n-1$, ne pourra jamais être égale au cycle $(1\ 2\ \dots\ n)$. Cela se voit aisément grâce à la représentation des factorisations minimales sous forme d'arbres que nous allons voir ci-dessous.



Représentation sous forme de graphe. À un k -uplet de transpositions (τ_1, \dots, τ_k) de \mathfrak{S}_n on associe un **graphe** à n sommets numérotés de 1 à n , noté $\mathcal{F}(\tau_1, \dots, \tau_k)$, en traçant une arête étiquetée i entre les sommets a et b si, et seulement si, $\tau_i = (a b)$ (voir Figure A.5 à gauche).

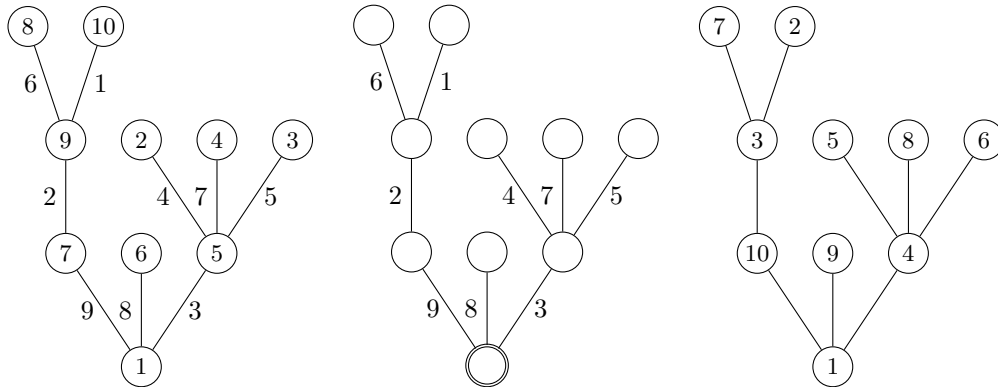


FIGURE A.5 – À gauche est représenté le graphe $\mathcal{F}(f_n)$ avec $n = 10$ associé à la factorisation minimale $f_n = ((9\ 10), (7\ 9), (1\ 5), (2\ 5), (3\ 5), (8\ 9), (4\ 5), (1\ 6), (1\ 7))$ de $(1 \dots 10)$. Au milieu est représenté $\mathcal{E}(f_n) \in \mathcal{C}'_n$. À droite est représenté $\mathcal{C}(f_n) \in \mathcal{C}_n$

La chose frappante à remarquer dans la Figure A.5 à gauche est que le graphe $\mathcal{F}(\tau_1, \dots, \tau_{n-1})$ est connexe et n'a pas de cycles. Autrement dit c'est un arbre. Ce fait est général pour les factorisations minimales (mais ne l'est pas pour les k -uplets quelconques de transpositions). En fait, $\mathcal{F}(\tau_1, \dots, \tau_k)$ est un arbre si, et seulement si, $k = n-1$ et $(\tau_1, \dots, \tau_k) \in \mathcal{F}_n$. Cela se voit même aisément : $\tau_k \circ \dots \circ \tau_1 = (1\ 2 \dots n)$ si et seulement si $\mathcal{F}(\tau_1, \dots, \tau_k)$ est connexe. Auquel cas, ce graphe, qui possède n sommets, contient au moins $n-1$ arêtes. Or le nombre d'arêtes de ce graphe est k qui correspond au nombre de transpositions. On en déduit donc que $k \geq n-1$ ce qui justifie finalement le qualificatif « minimales ». La fonction \mathcal{F} nous permet donc de transformer un élément de \mathcal{F}_n en un arbre à n sommets. Cependant \mathcal{F} ne constitue pas directement la bijection recherchée avec les arbres de Cayley. La première raison étant que l'image par \mathcal{F} n'est pas un arbre de Cayley mais plutôt un arbre étiqueté par les sommets et les arêtes, et aussi parce que \mathcal{F} n'est pas surjective dans l'ensemble de tels arbres étiquetés. En revanche il est vrai que \mathcal{F} est injective : on peut retrouver une factorisation minimale à partir de son image par \mathcal{F} . Cette fonction n'est pas surjective car il y a une condition sur les étiquettes des sommets, ces dernières ne peuvent pas se placer n'importe comment. En réalité, si l'on efface toutes ces étiquettes, sauf la 1, alors il n'y a qu'une seule manière de retrouver les autres de sorte que l'arbre étiqueté obtenu provienne bien d'une factorisation minimale. Plus précisément, on définit l'arbre **enraciné** $\mathcal{E}(\tau_1, \dots, \tau_{n-1})$ à partir de $\mathcal{F}(\tau_1, \dots, \tau_{n-1})$ en l'enracinant en le sommet étiqueté 1 et en y effaçant les étiquettes des sommets (voir Figure A.5 au milieu).

Proposition I.A.9 : Féray & Kortchemski [52]



La fonction \mathcal{E} constitue une bijection entre \mathcal{F}_n et l'ensemble \mathcal{C}'_n des arbres enracinés à $n-1$ arêtes étiquetées de 1 à $n-1$.

L'ensemble \mathcal{C}'_n ne correspond pas exactement à l'ensemble \mathcal{C}_n des arbres de Cayley à n sommets, puisque, dans le premier, les arbres sont enracinés et étiquetés par les arêtes, tandis que dans le second, il n'y a pas



de racine et ce sont les sommets qui sont étiquetés. En réalité les ensembles \mathcal{C}'_n et \mathcal{C}_n sont en bijection. Si $t \in \mathcal{C}_n$, on crée l'arbre $\alpha(t) \in \mathcal{C}'_n$ en enracinant t en 1, puis en tirant les étiquettes vers la racine (sauf 1 qui est effacé) pour terminer sur l'arête la plus proche, et enfin en ôtant 1 à toutes les étiquettes (voir Figure A.6). La fonction α est une bijection de \mathcal{C}'_n dans \mathcal{C}_n (sa bijection réciproque se décrit aisément). Pour une factorisation minimale $f_n \in \mathcal{F}_n$, on définit alors $\mathcal{C}(f_n) := \alpha \circ \mathcal{E}(f_n)$ (voir Figure A.5 à droite). La fonction \mathcal{C} forme finalement notre bijection entre \mathcal{F}_n et \mathcal{C}_n .

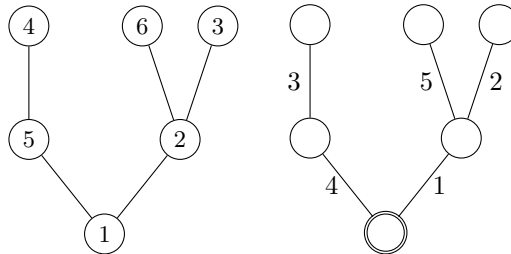


FIGURE A.6 – Exemple d'un arbre de Cayley $t \in \mathcal{C}_6$ à gauche transformé en $\alpha(t) \in \mathcal{C}'_6$ en tirant les étiquettes vers la racine et en retranchant 1.

Nous avons donc à notre disposition une bijection entre les factorisations minimales et les arbres de Cayley. L'étude des factorisations peut donc se ramener à celle de ces arbres. Si l'on pioche un élément de \mathcal{F}_n uniformément au hasard, alors son image par la bijection \mathcal{C} est aussi uniforme dans \mathcal{C}_n . Le comportement d'un arbre aléatoire uniforme dans \mathcal{C}_n est bien connu et s'inscrit dans le cadre plus général des arbres dit de Bienaymé, comme nous le verrons plus tard Section C.i. Une théorie bien développée sur le comportement asymptotique des arbres de Bienaymé, sous certaines conditions, permet de déduire celui des factorisations minimales lorsque n tend vers l'infini. Féray et Kortchemski [51] et [52] ont exploité à deux reprises la théorie des arbres de Bienaymé pour décrire certains aspects asymptotiques des factorisations minimales. La première démarche de Féray et Kortchemski consiste à étudier ce que l'on appelle les **laminations** associées aux factorisations minimales. Pour faire simple, une lamination est un ensemble de cordes du cercle unité qui ne se croisent pas. Cette démarche est donc plutôt géométrique et nous y reviendrons dans la Section D.iv. Pour l'instant, nous nous concentrons sur la deuxième démarche qui consiste à étudier, non pas les laminations, mais les **trajectoires** associées aux factorisations minimales.

Trajectoires d'une factorisation minimale. Développons ici la démarche de Féray et Kortchemski [52] qui consiste à étudier les trajectoires d'une factorisation minimale choisie uniformément au hasard $(\tau_1^{(n)}, \dots, \tau_{n-1}^{(n)}) \in \mathcal{F}_n$. Pour tout $i \in \llbracket 1, n \rrbracket$, la *trajectoire* de i est la fonction aléatoire

$$X_i^{(n)}(k) = \tau_k^{(n)} \circ \dots \circ \tau_1^{(n)}(i) \quad \forall k \in \llbracket 1, n-1 \rrbracket.$$

La fonction $X_i^{(n)}$ correspond à la trajectoire décrite par i au fur et à mesure qu'on lui applique les transpositions de la factorisation minimale. Féray et Kortchemski étudient la convergence de ces trajectoires quand n tend vers l'infini. Ils obtiennent que celles qui passent proche de 0 convergent conjointement vers des trajectoires limites [52, Théorème 1.1]. La technique utilisée pour obtenir ce résultat consiste à exploiter, comme nous l'avons dit ci-dessus, la théorie des arbres de Bienaymé. Grâce à cette dernière, les auteurs montrent que l'arbre étiqueté $\mathcal{E}(\tau_1^{(n)}, \dots, \tau_{n-1}^{(n)})$, après avoir divisé toutes les étiquettes par n , converge localement vers un arbre limite (nous verrons la notion de convergence locale Section D.i). L'enjeu de [52] est de voir si cet arbre

limite correspond aussi à une sorte de factorisation minimale infinie. Autrement dit, le but serait de pouvoir lire des trajectoires sur l'arbre limite en adaptant l'algorithme permettant de lire les trajectoires $X_i^{(n)}$ sur $\mathcal{E}(\tau_1^{(n)}, \dots, \tau_{n-1}^{(n)})$. Nous présentons ici un algorithme qui explique comment faire cette opération. Soit $t_n \in \mathcal{C}'_n$. Pour retrouver la factorisation minimale $f_n = \mathcal{E}^{-1}(t_n) \in \mathcal{F}_n$ associée à t_n , il faut appliquer l'algorithme suivant, qui provient de [52, Section 2.3] et que nous notons (\mathcal{A}) .

- Étape 1 : attribuer l'étiquette 1 à la racine de t_n .
- Étape 2 : partir du sommet étiqueté 1 et suivre le plus long chemin croissant d'arêtes en choisissant à chaque intersection la plus petite arête possible. Puis donner 2 comme étiquette au sommet à l'extrémité du chemin.
- Étape 3 : répéter la deuxième étape mais en partant cette fois du sommet 2 pour trouver le sommet 3.
- Étape 4 à n : répéter encore jusqu'à avoir étiqueté tous les sommets de t_n de 1 à n .

Une fois l'algorithme (\mathcal{A}) terminé, nous obtenons un arbre étiqueté par les sommets et les arêtes qui correspond à $\mathcal{F}(f_n)$. La i -ième transposition de f_n est alors $\tau_i = (a b)$ où a et b sont les sommets adjacents à l'arête i . De plus la trajectoire X_i correspond au chemin parcouru par l'algorithme partant du sommet i (voir dans la Figure A.5 comment l'algorithme (\mathcal{A}) appliqué à l'arbre du milieu permet de retrouver les étiquettes de l'arbre à gauche). Féray et Kortchemski ont adapté l'algorithme ci-dessus pour l'appliquer à l'arbre limite.

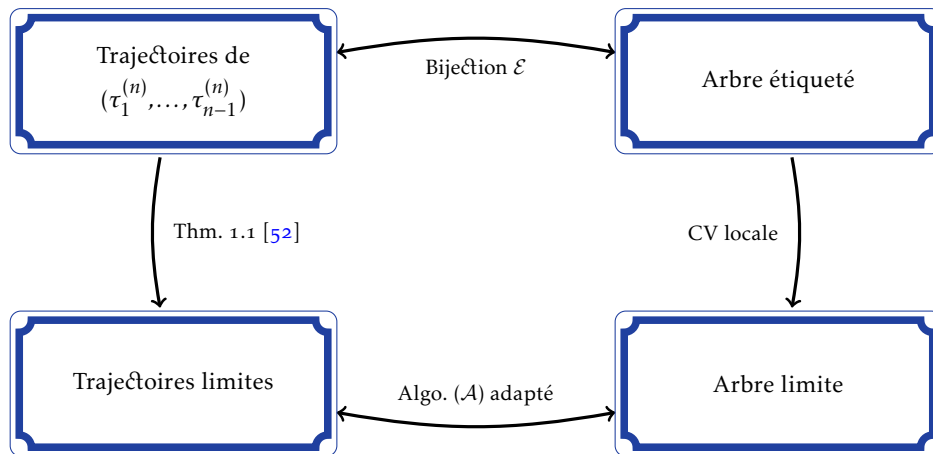


FIGURE A.7 – Stratégie de [52] pour étudier les trajectoires d'une factorisation minimale choisie uniformément au hasard.

Comme conséquence du Théorème 1.1 de [52] les auteurs obtiennent, entre autres,

Théorème I.A.10 : Féray & Kortchemski [52]

Pour tout $i, j \geq 1$, on a la convergence

$$\mathbb{P}\left(\mathbb{T}_1^{(n)} = i, \mathbb{T}_2^{(n)} = j\right) \xrightarrow{n \rightarrow \infty} e^{-2} \left[\frac{i+j-2}{(i+j-1)!} + \frac{i+j-1}{i!j!} - \frac{i+j-1}{(i+j)!} \right].$$

où $\mathbb{T}_k^{(n)} = \#\{1 \leq \ell \leq n-1 : \tau_\ell^{(n)}(k) \neq k\}$ correspond au nombre de fois que k apparaît dans une factorisation minimale $(\tau_1^{(n)}, \dots, \tau_{n-1}^{(n)})$ tirée uniformément au hasard dans \mathcal{F}_n .

Étudier la limite locale d'arbres aléatoires est une approche très probabiliste. Il s'avère qu'avec un point de vue purement combinatoire il est possible de retrouver I.A.10 et même de généraliser la formule pour tout n (nous verrons Section B.iii comment, au contraire, un point de vue probabiliste permettra de généraliser un résultat combinatoire). Plus précisément

Théorème I.A.11 : B. [14]



Pour tout $i, j \geq 1$ et $n \geq i + j$

$$\mathbb{P}\left(\mathbb{T}_1^{(n)} = i, \mathbb{T}_2^{(n)} = j\right) = \frac{n!(n-1)^{n-i-j-1}}{(n-i-j)!(n+1)^{n-1}} \left[\frac{(i+j-2)(n-1)}{(i+j-1)!(n-i-j+1)} + \frac{i+j-1}{i!j!} - \frac{i+j-1}{(i+j)!} \right].$$

Le résultat précédent se déduit du Théorème I.A.6 et de l'étude d'une autre fonction génératrice :

$$G_n(x, y, z) = n^{n-2} \mathbb{E}\left[x^{\mathbb{T}_1^{(n)}} y^{\mathbb{T}_2^{(n)}} z^{\mathbb{M}_1^{(n)}}\right]$$

où $\mathbb{M}_i^{(n)} = \#\{1 \leq k \leq n-1 : \tau_{k-1}^{(n)} \circ \dots \circ \tau_1^{(n)}(i) \neq \tau_k^{(n)} \circ \dots \circ \tau_1^{(n)}(i)\}$ correspond au nombre de transpositions qui affectent la trajectoire de i . Grâce à la bijection \mathcal{C} entre \mathcal{F}_n et \mathcal{C}_n , la fonction génératrice G_n peut se réécrire en termes d'arbres de Cayley :

$$G_n(x, y, z) = \sum_{t \in \mathcal{C}_n} x^{\deg_1(t)} y^{\deg'_2(t)} z^{|L_1(t)|}$$

où $L_1(t)$ correspond au chemin dans t partant de 1 décrit Section A.i et $\deg'_2(t)$ correspond au degré du dernier sommet du chemin $L_1(t)$. La quantité $|L_1|$ est donc l'équivalent, pour les arbres de Cayley, de la quantité $\mathbb{M}_1^{(n)}$, pour les factorisations minimales. La différence entre G_n et F_n est que \deg_2 est devenu \deg'_2 dans G_n . La preuve de Théorème I.A.11 consiste à calculer la série exponentielle associée à G_n et par identification des coefficients à en déduire la loi jointe de $\mathbb{T}_1^{(n)}$ et $\mathbb{T}_2^{(n)}$. L'étude de la série exponentielle se fait grâce à diverses techniques combinatoires et grâce à la formule explicite de F_n trouvée au Théorème I.A.6. Une certaine symétrie du problème est également exploitée afin de dire que la fonction génératrice G_n est invariante par permutation de x et y .

Remarque

Deux questions peuvent être posées :

1. Pourquoi, dans le Théorème I.A.11, se contenter de la loi de $(\mathbb{T}_1^{(n)}, \mathbb{T}_2^{(n)})$ et ne pas regarder $(\mathbb{T}_1^{(n)}, \mathbb{T}_2^{(n)}, \mathbb{M}_1^{(n)})$? Il serait possible en pratique, en extrayant les coefficients de G_n , d'obtenir la loi jointe de ce triplet. Cependant il semblerait que la formule ne soit pas aussi simple et concise que celle du Théorème I.A.11. C'est pourquoi nous nous sommes contenté de cette dernière.
2. Pourquoi intégrer le paramètre $\mathbb{M}_1^{(n)}$ si seule l'étude du couple $(\mathbb{T}_1^{(n)}, \mathbb{T}_2^{(n)})$ nous intéresse ? L'équivalent de $\mathbb{T}_2^{(n)}$, pour les arbres, est \deg'_2 qui correspond au degré du sommet final du chemin L_1 . On comprend donc que la connaissance de ce chemin est indispensable dans l'étude de \deg'_2 . Nous sommes donc face à une situation, typique en mathématiques, où rajouter un paramètre simplifie le problème au lieu de le complexifier.



B Arborescences ordonnées

Dans cette section, nous introduisons la notion d'arbre ordonné et d'arbre plan. Jusqu'à présent, la manière avec laquelle les arbres étaient représentés dans le plan n'était pas importante (cf Figure A.1). Dans le cas des arbres plans, cela va le devenir.

B.i Prolégomènes

Définition I.B.1 : Arbre ordonné

Un *arbre ordonné* est la donnée d'un arbre $T = (S, A)$ et, pour tout sommet $u \in S$, d'un ordre \leq_u sur $V(u)$, où $V(u)$ désigne l'ensemble des voisins de u .

La *forme* d'un arbre ordonné est l'arbre sous-jacent, au sens de la définition I.A.1, c'est-à-dire où l'ordre est ignoré.

Définition I.B.2 : Arbre plan

Un *arbre plan* est un arbre ordonné $T = (S, A)$ tel que pour tout $u \in S$, l'ensemble ordonné $(V(u), \leq_u)$ est isomorphe à une partie de \mathbb{R} muni de l'ordre canonique. On dira aussi dans ce cas que l'ordre de T est *planaire*.

Remarque

La définition que nous donnons ici d'arbre plan n'est pas usuelle dans la communauté probabiliste. En effet, il est plus fréquent de faire appel, par exemple, au formalisme de Neveu [86]. Les raisons qui nous poussent à utiliser la Définition I.B.2 plutôt que le formalisme de Neveu est que ce dernier n'autorise pas les sommets à avoir un nombre indénombrable de voisins (même si en pratique cette restriction ne nous affectera pas dans nos applications). Notre dernière remarque sera la suivante : il ne faut pas confondre arbre plan et arbre planaire. En effet, un graphe planaire est un graphe pouvant être dessiné dans le plan sans que les arêtes ne se croisent. Cela est donc bien différent du concept d'arbre plan.

Intuitivement, dire que, pour tout sommet u , $(V(u), \leq_u)$ est isomorphe à une partie de \mathbb{R} revient à dire que l'arbre T peut se dessiner dans le plan de sorte que l'ordre naturel de rotation dans le plan correspond à l'ordre \leq_u . Nous détaillons cela dans le prochain paragraphe.

Dessiner un arbre plan sur un plan. Lorsque l'on dessinera un arbre plan sur un plan, tel qu'une feuille de papier ou encore un écran d'ordinateur, on s'appliquera à ce que l'ordre des voisins d'un sommet u soit donné par la rotation autour de u en partant du vecteur de la base canonique $(1, 0)$. L'ordre des sommets voisins de u correspondra à leurs instants d'apparition dans cette rotation. De cette manière, on pourra se dire que l'on a *plongé* l'arbre dans le plan. Il faut faire attention au fait suivant : ce n'est pas parce qu'un arbre est dessiné sur le plan que celui est nécessairement un arbre plan (ou même ordonné). Dans tout ce document, nous allons fréquemment dessiner des arbres non-ordonnés dans le plan. Il faudra prêter attention au contexte afin de déterminer si l'arbre est muni d'un ordre (planaire) ou non. Voir Figures B.1 et B.2.



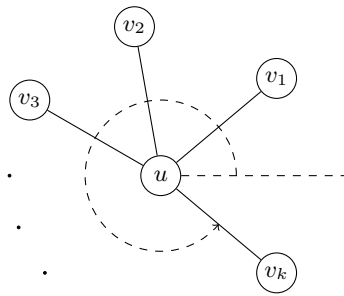


FIGURE B.1 – Arbre plan plongé dans le plan. Ici $v_1 \leq_u v_2 \leq_u v_3 \leq_u \dots \leq_u v_k$.

Lorsqu’un arbre plan est enraciné, l’ordre autour d’un sommet peut s’interpréter comme l’âge des individus : le plus vieux est le parent, le plus vieil enfant est l’aîné, le plus jeune enfant est le cadet, etc. Le champ lexical familial est très pratique dans l’étude des arbres plans (et des arbres en général).

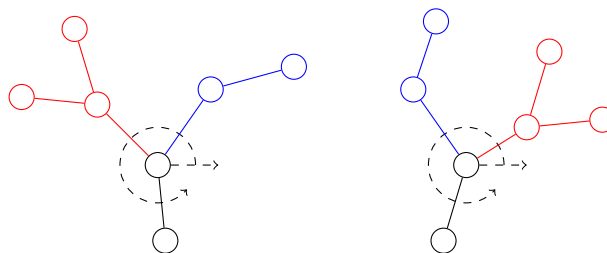


FIGURE B.2 – Deux arbres plans dessinés dans le plan. Ces deux arbres ont la même forme mais l’ordre autour du sommet central est différent (les parties rouges et bleues sont échangées), ils constituent donc deux arbres plans **non isomorphes**.

Comme pour les arbres de Cayley à n sommets, il est possible de compter les arbres plans à n sommets à isomorphisme près. Plus précisément on considère que deux arbres plans sont isomorphes si on peut passer de l’un à l’autre par simple renommage des sommets et si les ordres, dans les deux arbres, autour d’un même sommet, sont isomorphes. Dans la suite on considèrera tacitement que l’on considère les arbres plans à isomorphisme près.

Théorème I.B.3 : Dénombrement des arbres plans



Le nombre d’arbres plans enracinés à $n + 1$ sommets, à isomorphisme près, correspond au n -ième nombre de Catalan

$$C_n = \frac{1}{n + 1} \binom{2n}{n}.$$

Ordre canonique d’un arbre enraciné étiqueté. Dans les Sections B.iii et B.iv nous aurons besoin d’attribuer un ordre planaire à un arbre étiqueté par les sommets ou les arêtes. De manière générale, si T est un arbre enraciné, étiqueté par les sommets et que, pour tout sommet u de T , les enfants de u sont étiquetés par des nombres réels deux à deux distincts, on pourra attribuer un ordre canonique autour de u . Cet ordre s’obtiendra en déclarant que le plus grand élément de \leq_u est le père de u et les enfants de u seront ordonnés selon l’ordre **inverse** de leurs étiquettes. Voir Figure B.3 à gauche. De manière similaire, si T est un arbre enraciné, étiqueté

par les arêtes et que, pour tout sommet u de T , les arêtes sortant de u sont étiquetées par des nombres réels deux à deux distincts, on pourra attribuer un ordre canonique autour de u . Cet ordre s'obtiendra en déclarant que le père de u sera le plus grand élément de \leq_u et les enfants de u seront ordonnés selon l'ordre **inverse** des étiquettes des arêtes les reliant à u . Voir Figure B.3 à droite.

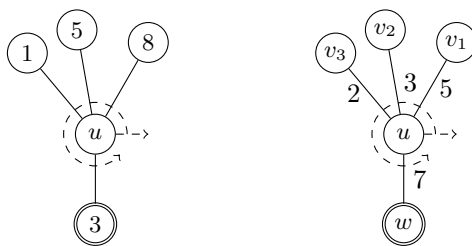


FIGURE B.3 – À gauche, un arbre étiqueté par les sommets avec son ordre planaire canonique. Notons $s(i)$ le sommet étiqueté i . Ici $s(3)$ est le parent de u et $s(1)$, $s(5)$ et $s(8)$ sont les enfants de u . Pour l'ordre planaire canonique on a $s(8) \leq_u s(5) \leq_u s(1) \leq_u s(3)$. À droite, un arbre étiqueté par les arêtes avec son ordre planaire canonique. Ici w est le parent de u et v_1 , v_2 et v_3 sont les enfants de u . Pour l'ordre planaire canonique on a $v_1 \leq_u v_2 \leq_u v_3 \leq_u w$.

B.ii Parcourir un arbre plan fini et enraciné

Soit T un arbre plan fini et enraciné. *Parcourir* T signifie explorer les sommets de T un à un en suivant un algorithme déterminé jusqu'à les avoir tous explorés. Plus précisément on tient à jour une liste de sommets explorés parmi l'ensemble des sommets de T . À chaque étape $k \geq 0$ du parcours on note E_k la liste des sommets explorés à l'étape k . On initialise le parcours avec E_0 qui est constitué uniquement du sommet racine. Ensuite on explore les sommets de T en fonction de l'algorithme de parcours choisi. Il en existe plusieurs, nous utiliserons principalement le parcours en profondeur et le parcours en largeur que nous décrivons ci-dessous.

Parcours en profondeur. Avant de décrire le parcours en profondeur, nous définissons la notion d'ordre lexicographique. Soit T un arbre plan fini et enraciné. L'ordre lexicographique sur T est un ordre total sur les sommets de T , noté \leq_{lex} , et défini de la manière suivante. Soit u, v des sommets différents de T , on note w leur plus proche ancêtre commun (ici « ancêtre » est à prendre au sens large, c'est-à-dire qu'un sommet est son propre ancêtre), u' le fils de w qui est ancêtre de u et v' le fils de w qui est ancêtre de v . On note alors $v \leq_{\text{lex}} u$ si $u' \leq_w v'$ et $u \leq_{\text{lex}} v$ sinon. Le *parcours en profondeur* de T consiste alors à explorer les sommets de T en suivant leur rang dans l'ordre lexicographique, voir Figure B.4 à gauche. Intuitivement, le parcours en profondeur consiste à explorer les sommets en priorisant les sommets les plus haut puis les plus à gauche parmi les sommets adjacents aux sommets déjà explorés.

Parcours en largeur. De manière semblable au parcours en profondeur, on définit d'abord un ordre total, noté \leq_{lrg} , sur les sommets d'un arbre plan fini et enraciné T . Soit u, v deux sommets différents de T , on note alors $u \leq_{\text{lrg}} v$ si la hauteur de u est plus faible que celle de v ou si u et v ont la même hauteur mais $u \leq_{\text{lex}} v$ (où \leq_{lex} désigne l'ordre lexicographique décrit plus haut). Sinon on note $v \leq_{\text{lrg}} u$. Le *parcours en largeur* de T consiste alors à explorer les sommets de T en suivant leur rang pour l'ordre \leq_{lrg} , voir Figure B.4 à droite. Intuitivement, le parcours en largeur consiste à explorer les sommets en priorisant les sommets les plus bas puis les plus à gauche parmi les sommets adjacents aux sommets déjà explorés.

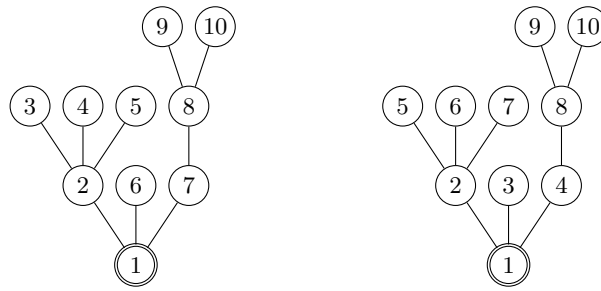


FIGURE B.4 – Représentation d’un arbre plan fini et enraciné T . À gauche les sommets de T sont étiquetés selon leur rang dans le parcours en profondeur. À droite ils sont étiquetés selon leur rang dans le parcours en largeur.

Marche de Łukasiewicz . La marche de Łukasiewicz est une manière de coder un arbre plan fini et enraciné à travers une marche. Dans le cadre probabiliste cela permet de transposer l’étude d’arbres aléatoires à l’étude de marches aléatoires pour lesquelles une théorie bien développée existe. Le codage par la marche de Łukasiewicz passe par un parcours en profondeur de l’arbre. Plus précisément, soit T un arbre ordonné et enraciné avec $n \in \mathbb{N}^*$ sommets. Pour tout sommet u de T on note k_u le nombre d’enfants de u (donc $k_u = \text{deg}(u) - 1$ si u n’est pas racine et $k_u = \text{deg}(u)$ si u est racine). Notons u_1, u_2, \dots, u_n les sommets de T rangés dans l’ordre lexicographique (donc u_1 est la racine de T). On définit la marche de Łukasiewicz $(S_i)_{0 \leq i \leq n}$ de la manière suivante : on initialise avec $S_0 := 0$ et pour tout $i \in \llbracket 1, n \rrbracket$, on pose $S_i := S_{i-1} + k_{u_i} - 1$ (voir Figure B.5). À noter que la marche $(S_i)_i$ ne fait que des pas à valeurs dans $\mathbb{N} \cup \{-1\}$ et qu’elle touche -1 pour la première fois au temps n . La propriété centrale de la marche de Łukasiewicz est qu’elle code parfaitement l’arbre associé, autrement dit, il est possible de retrouver sans ambiguïté l’arbre à partir de sa marche de Łukasiewicz seule. La marche de Łukasiewicz peut s’adapter au parcours en largeur : le procédé reste le même mais les sommets sont ordonnés selon \leq_{lr} et non plus selon l’ordre lexicographique \leq_{lex} . Ce codage modifié reste bijectif et les propriétés essentielles se conservent. Nous utiliserons dans cette thèse ces deux versions de la marche de Łukasiewicz , la version parcours en profondeur lors de l’étude des factorisations monotones et le parcours en largeur lors de l’étude des tuples de stationnement.

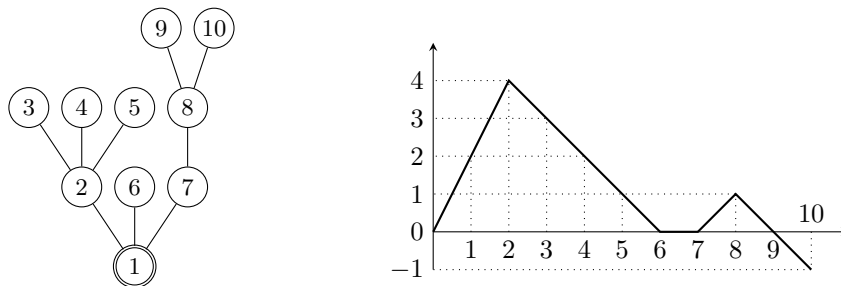


FIGURE B.5 – À gauche, un arbre plan enraciné à 10 sommets étiquetés selon leur rang dans l’ordre lexicographique. À droite, la marche de Łukasiewicz associée à l’arbre de gauche.

B.iii Application aux tuples de Stationnement

Nous avons mentionné, Section A.ii, que les factorisations minimales étaient liées à un autre objet combinatoire appelé « tuple de stationnement ». Cet objet est le sujet de cette section. Nous verrons, qu'une fois de plus, une approche bijective à travers les arbres nous permettra d'étudier ce nouvel objet et d'obtenir de nouveaux résultats. Définissons d'abord les tuples de stationnement. Un n -uplet de stationnement peut se voir de la manière suivante. Imaginons que n voitures souhaitent se garer successivement dans un parking à n places numérotées de 1 à n . Les places sont situées sur une seule ligne et ordonnées de gauche à droite. Chaque conducteur a sa place préférée, notons $\pi_n(i)$ la place préférée du conducteur de la i -ième voiture. À l'étape i , le conducteur de la i -ième voiture va vouloir se garer à sa place préférée $\pi_n(i)$. Si cette place est libre lorsqu'il arrive, il s'y gare et la $(i+1)$ -ième voiture arrive sur le parking. Sinon il va se décaler d'une place et chercher à se garer à la place $\pi_n(i)+1$. S'il peut s'y garer, il le fait et on passe à la voiture suivante, sinon il se décale une nouvelle fois vers la droite et ainsi de suite. Si le conducteur de la i -ième voiture finit par arriver à la place numéro n sans avoir réussi à se garer et si cette place est encore occupée, alors il quitte le parking. Un tuple de stationnement est donc la donnée de n places préférées $\pi_n(1), \dots, \pi_n(n)$ telles que chaque voiture pourra se garer (voir Figure B.6). Il s'avère que ce critère de stationnement est équivalent à celui donné dans la Définition I.B.4 ci-dessous, plus formelle et succincte.

Définition I.B.4 : Tuple de Stationnement

Un n -uplet de stationnement est un n -uplet $\pi_n = (\pi_n(1), \dots, \pi_n(n))$ tel que, si $\pi'_n(1) \leq \dots \leq \pi'_n(n)$ est le réordonnement croissant de $\pi_n(1), \dots, \pi_n(n)$, alors $\pi'_n(i) \leq i$ pour tout $i \in \llbracket 1, n \rrbracket$. On note \mathcal{S}_n l'ensemble des n -uplets de stationnement.

Par exemple,

- Le n -uplet $(1, 1, \dots, 1)$ est toujours un n -uplet de stationnement.
- Le 9-uplet $(4, 6, 2, 7, 6, 2, 1, 1, 4)$ est de stationnement.
- le triplet $(2, 2, 3)$ n'est pas de stationnement car la troisième voiture n'arrivera pas à se garer.

Remarque

Cette métaphore des voitures justifie le nom de « tuple de stationnement ». N'ayant pas réussi à trouver de référence satisfaisante en français traitant du sujet, nous nous sommes permis de traduire le terme « parking function », apparaissant dans la plupart des références anglophones, par « tuple de stationnement ». Il nous a semblé plus adapté de parler de « tuple de stationnement » plutôt que de « fonction de stationnement » étant donné l'origine informatique de cet objet, et aussi car la notation en terme de tuple semble plus courante que celle en terme de fonction y compris dans les références anglophones. De plus, un avantage du terme « tuple » sur le terme « fonction » est que l'on peut facilement dire un « n -uplet de stationnement » au lieu de dire « fonction de parking de taille n » ou encore de devoir introduire une formulation du style « n -fonction de stationnement » (qui ne sonne pas très naturel en français contrairement à l'anglais « n -parking function »).

Motivations et bibliographie. Konheim et Weiss [75] ont introduit les tuples de stationnement alors qu'ils étudiaient des fonctions de hachage. Ils ont montré qu'il y a $(n + 1)^{n-1}$ n -uplets de stationnement. Depuis, les tuples de stationnement sont devenus un sujet d'étude en combinatoire, théorie des probabilités, théorie des groupes et, bien-sûr, en informatique. Plus précisément, les tuples de stationnement sont liés à la théorie énumérative des arbres et forêts [35], aux processus de coalescences [34, 28], à l'analyse de partitions d'ensembles [94], d'arrangements d'hyperplans [95, 92], polytopes [96, 36] et groupe « tas de sable » [37]. L'étude des propriétés probabilistes des tuples de stationnement a récemment attiré de l'intérêt dans la communauté [45, 73, 104]. Voir par exemple [103] pour une revue assez large du sujet. Comme nous l'avons dit, la raison pour laquelle nous avons étudié les tuples de stationnement, est qu'ils sont en connexion avec les factorisations minimales du cycle.

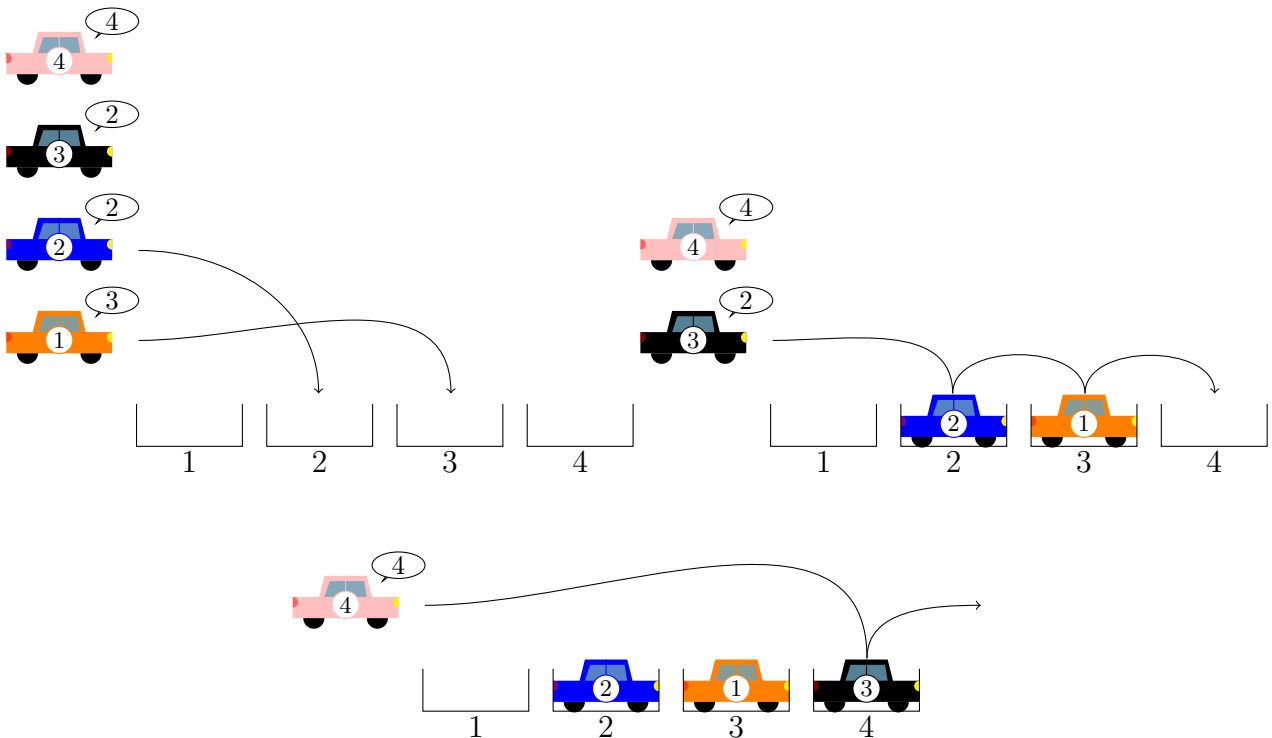


FIGURE B.6 – Quatre voitures numérotées de 1 à 4 cherchent à se garer l’une après l’autre sur un parking à quatre places, également numérotées de 1 à 4. la voiture 1 souhaite se garer à la place numéro 3, la voiture 2 à la place 2, etc. Les voitures 1 et 2 parviendront à se garer sur leur place souhaitée car elle sera libre. La voiture 3 devra se décaler à deux reprises avant de trouver une place libre. La voiture 4, quant à elle, ne parviendra pas à se garer et devra quitter le parking. Le quadruplet $(3, 2, 2, 4)$ n’est donc pas dans S_4 .

Bijection avec les factorisations minimales. Détaillons dans ce paragraphe le lien entre tuples de stationnement et factorisations minimales. Biane [20] démontre que l’application

$$((a_1 \ b_1), \dots, (a_{n-1} \ b_{n-1})) \rightarrow (a_{n-1}, a_{n-2}, \dots, a_1)$$

où $a_i \leq b_i$, est une bijection entre \mathcal{F}_n et \mathcal{S}_n . Cela confirme bien que le cardinal de \mathcal{S}_n est $(n + 1)^{n-1}$ et incite à penser qu’il peut exister une bijection directe avec les arbres de Cayley. Nous le confirmerons dans le paragraphe suivant. À travers cette bijection, le Théorème I.A.11 permet de décrire la loi du nombre de fois que 1 et 2 apparaissent dans un n -uplet de stationnement choisi uniformément au hasard.

Bijection avec les arbres. Dans la littérature, plusieurs bijections ont été décrites entre les tuples de stationnement et les arbres. Celle que nous utiliserons vient de Chassaing et Marckert [35]. Cette bijection utilise le parcours en largeur vu dans la Section B.ii. Plus précisément, soit $t \in \mathcal{C}_{n+1}$ un arbre de Cayley à $n+1$ sommets. Avant de pouvoir parcourir l'arbre t , il faut d'abord le munir d'une structure d'arbre plan enraciné. Pour ce faire, nous enracinons t en son sommet numéroté 1. Ensuite, nous ordonnons chaque enfant d'un même sommet dans l'ordre décroissant de leurs étiquettes comme décrit Section B.i. Pour tout $i \in \llbracket 1, n \rrbracket$, on définit $r(i, t)$ le rang du **parent** du sommet i pour l'ordre \leq_{lrg} (qui correspond à l'ordre du parcours en largeur) dans t . La bijection de Chassaing et Marckert est décrite dans le théorème suivant (voir Figure B.7).

Théorème I.B.5 : Chassaing & Marckert [35]



La fonction

$$t \mapsto (r(2, t), \dots, r(n+1, t)) \quad (\text{B.1})$$

est une bijection entre \mathcal{C}_{n+1} et \mathcal{S}_n .

Remarque

Chassaing et Louchard [34] ont décrit une bijection semblable en utilisant, ce qu'ils appellent, l'ordre standard au lieu de l'ordre \leq_{lrg} .

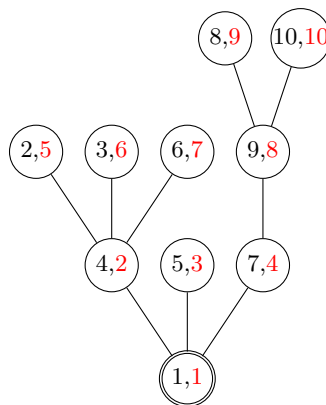


FIGURE B.7 – Exemple d'un arbre de Cayley t à 10 sommets enraciné et muni de son ordre planaire. L'étiquette de chaque sommet est représentée en noir à gauche et son rang pour l'ordre \leq_{lrg} du parcours en largeur est représenté en rouge à droite. Il est à noter que les enfants d'un même sommet ont leurs étiquettes en noir qui croissent de gauche à droite car c'est de cette manière que l'arbre a été plongé dans le plan. Par exemple, ici $r(6, t) = 2$. Le 9-uplet de stationnement associé à cet arbre par (B.1) est $(2, 2, 1, 1, 2, 1, 8, 4, 8)$.

La bijection (B.1) sera la première pierre de notre étude des tuples de stationnement. Cependant, pour aller plus loin, nous devons attendre de voir la notion d'arbre de Bienaymé Section C.i. En résumé, lorsque l'on choisit un n -uplet de stationnement uniformément au hasard, la marche de Łukasiewicz de l'arbre associé par (B.1) est une marche aléatoire de loi déterminée. En étudiant cette marche aléatoire nous pourrions déterminer certaines propriétés du n -uplet de stationnement. Nous verrons cela en détails Section C.ii après avoir vu la notion d'arbre de Bienaymé. En attendant, passons à une autre application des arbres plans qui est l'étude des



factorisations **monotones**.

B.iv Application aux factorisations monotones du cycle

Les factorisations monotones sont des factorisations minimales satisfaisant une condition supplémentaire. L'un des objectifs principaux de cette thèse était de transposer certains résultats des factorisations minimales aléatoires aux factorisations monotones aléatoires.

Définition I.B.6 : Factorisations monotones du cycle

Une *factorisation décroissante* (resp. *croissante*) de $(1\ 2\ \dots\ n)$ est une factorisation minimale $(\tau_1, \dots, \tau_{n-1})$ telle que, si $\tau_i = (a_i\ b_i)$ avec $a_i < b_i$, alors les $(a_i)_{1 \leq i \leq n-1}$ sont dans l'ordre décroissant (resp. croissant). On note \mathcal{F}_n^\downarrow (resp. \mathcal{F}_n^\uparrow) l'ensemble des factorisations minimales décroissantes (resp. croissantes).

Par exemple, le 9-uplet $((8\ 9), (8\ 10), (7\ 8), (2\ 3), (2\ 4), (2\ 5), (1\ 2), (1\ 6), (1\ 7))$ est une factorisation décroissante de $(1\ \dots\ 10)$ et $((1\ 4), (1\ 6), (2\ 4), (3\ 4), (5\ 6))$ est une factorisation croissante de $(1\ \dots\ 6)$.

Bibliographie et motivation. Les factorisations décroissantes (aussi appelées *factorisations primitives*) ont été étudiées par Gewurz & Merola [57] qui ont montré qu'elles étaient comptées par les nombres de Catalan. Deux bijections sont mentionnées dans [57] : la première

$$((a_1\ b_1), \dots, (a_{n-1}\ b_{n-1})) \rightarrow (a_{n-1}, a_{n-2}, \dots, a_1)$$

qui correspond à celle déjà mentionnée plus haut [20], est une bijection avec les $(n-1)$ -uplets de stationnement croissants, c'est-à-dire les $(n-1)$ -uplets de stationnement $(\pi_1, \dots, \pi_{n-1})$ tels que $\pi_1 \leq \dots \leq \pi_{n-1}$. La deuxième

$$((a_1\ b_1), \dots, (a_{n-1}\ b_{n-1})) \rightarrow (b_{n-1}, b_{n-2}, \dots, b_1)$$

est une bijection avec les permutations 231 -évitantes de taille $n-1$, c'est-à-dire, les permutations $\rho \in \mathfrak{S}_{n-1}$ telles qu'il n'existe pas d'indices $i < j < k$ vérifiant $\rho(k) < \rho(i) < \rho(j)$. Les factorisations croissantes apparaissent dans [64] et par le calcul de la fonction génératrice du Théorème 5.2, elles sont aussi comptées par les nombres de Catalan.

Bijection avec les arbres. Comme nous l'avons vu Section A.ii il est possible de coder toute l'information d'une factorisation minimale f_n de \mathcal{F}_n à travers un arbre $\mathcal{E}(f_n)$ enraciné de \mathfrak{C}'_n avec $n-1$ arêtes étiquetées de 1 à $n-1$. Il s'avère qu'en imposant f_n monotone (c'est-à-dire croissante ou décroissante), les étiquettes de $\mathcal{E}(f_n)$ deviennent **presque** redondantes. Plus précisément, si l'on attribue l'ordre planaire canonique à $\mathcal{E}(f_n)$, tel que décrit Section B.i, alors l'ordre planaire sur $\mathcal{E}(f_n)$ est suffisant pour retrouver f_n . Les étiquettes précises des arêtes ne sont donc pas nécessaires du moment que l'on connaît leur ordre parmi les enfants de chaque sommet. Nous avons donc une bijection entre $\mathcal{F}_n^\downarrow, \mathcal{F}_n^\uparrow$ et l'ensemble des arbres plans enracinés à n sommets. À noter que ce dernier a pour cardinal le n -ième nombre de Catalan ce qui corrobore les résultats sus-mentionnés. Plus formellement, appelons $\mathcal{P}(f_n)$ l'arbre $\mathcal{E}(f_n)$ muni de son ordre planaire canonique à qui on a effacé les étiquettes. Ainsi $\mathcal{P}(f_n)$ est un arbre plan enraciné à n sommets (sans aucune étiquette, voir Figure B.8).

Théorème I.B.7 : B. [16]

Pour tout $n \geq 1$, la fonction \mathcal{P} est une bijection entre \mathcal{F}_n^\downarrow (resp. \mathcal{F}_n^\uparrow) et l'ensemble des arbres plans enracinés à n sommets.

La démonstration de I.B.7 tient dans la description de la bijection réciproque à \mathcal{P} . Plus précisément, on explique comment retrouver les étiquettes des arêtes seulement à partir de l'ordre planaire. Cette description est assez délicate surtout dans le cas des factorisations croissantes. Se référer aux preuves des Propositions V.B.3 et V.B.5 afin d'avoir le procédé précis avec lequel retrouver les étiquettes. Cette bijection nous permettra d'étudier les laminations, que nous avons déjà évoqué, associées aux factorisations décroissantes dans la Section D.iv. En effet la bijection rend équivalente l'étude d'une factorisation décroissante aléatoire et celle d'un arbre plan aléatoire. Pour permettre cette dernière, comme dans le cas des tuples de stationnement, nous aurons besoin de quelques outils généraux sur les arbres de Bienaymé que nous verrons dans la prochaine section. Avant de passer à celle-ci, mentionnons des applications directes de notre bijection.

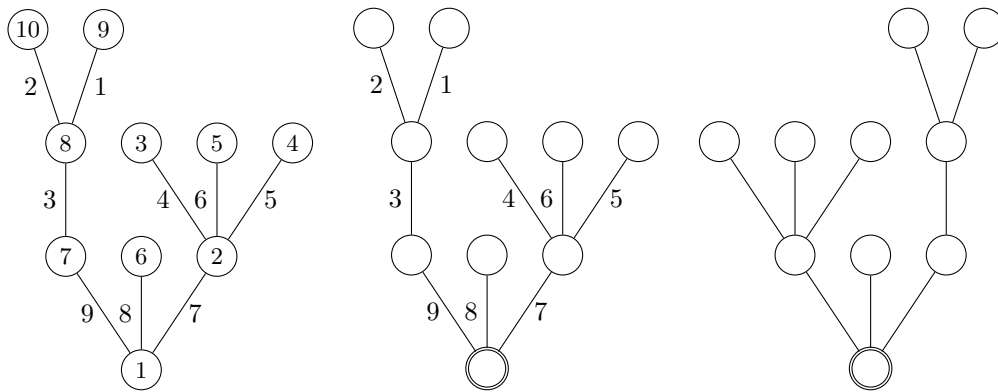


FIGURE B.8 – À gauche est représenté l'arbre $\mathcal{F}(f_n)$ avec $n = 10$ associé à la factorisation décroissante $f_n = ((8\ 9), (8\ 10), (7\ 8), (2\ 3), (2\ 4), (2\ 5), (1\ 2), (1\ 6), (1\ 7))$ de $(1 \cdots 10)$. Au milieu est représenté $\mathcal{E}(f_n) \in \mathcal{C}'_n$. À droite est représenté l'arbre plan enraciné $\mathcal{P}(f_n)$.

Conséquences de la bijection entre arbres plans et factorisations monotones.**Proposition I.B.8 : B. [16]**

Soit $(\tau_1, \dots, \tau_{n-1})$ une factorisation décroissante de taille n choisie uniformément au hasard avec $\tau_i = (a_i\ b_i)$ et $a_i < b_i$. Alors

1. $\{b_1, \dots, b_{n-1}\} = \{2, 3, \dots, n\}$ presque sûrement.
2. $\mathbb{P}(\#\{a_1, \dots, a_{n-1}\} = k) = \frac{n+1}{n-1} \binom{n-1}{k} \binom{n-1}{n-k} \binom{2n}{n}^{-1}$ pour tout $k \in \{1, \dots, n-1\}$.
3. La convergence

$$\frac{\#\{a_1, \dots, a_{n-1}\} - n/2}{\sqrt{n}} \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1/8)$$

tient en loi où $\mathcal{N}(0, 1/8)$ désigne une loi normale centrée de variance 1/8.



Remarque

Le premier point de la Proposition **I.B.8** a déjà été montré dans [57]. Comme sus-mentionné, les auteurs montrent même que (b_1, \dots, b_{n-1}) est une permutation 231-évitante.

Proposition I.B.9 : B. [16]

Soit $(\tau_1, \dots, \tau_{n-1})$ une factorisation croissante de taille n choisie uniformément au hasard avec $\tau_i = (a_i \ b_i)$ et $a_i < b_i$. Alors

1. Presque sûrement, les ensembles $\{a_1, \dots, a_{n-1}\}$ et $\{b_1, \dots, b_{n-1}\}$ forment une partition de $\{1, 2, 3, \dots, n\}$.
2. Les points 2 et 3 de la Proposition **I.B.8** sont toujours valables pour $\#\{a_1, \dots, a_{n-1}\}$.

Remarque

En fait la bijection de Biane [20], est aussi une bijection entre les tuples de stationnement décroissants et les factorisations croissantes. De plus on obtient une bijection sur les tuples de stationnement par une simple inversion de l'ordre des éléments. Ainsi le point 2 de la Proposition **I.B.9** se déduit directement de la Proposition **I.B.8**.



C Arborescences aléatoires

C.i Arbres de Bienaymé

Dans cette section, nous introduisons finalement les arbres de Bienaymé. Des résultats généraux sur ces arbres permettront de poursuivre notre étude des tuples de stationnement et des factorisations décroissantes. Les arbres de Bienaymé constituent l'un des modèles les plus simples d'arbres aléatoires et l'un des premiers modèles d'évolution de population. Ils peuvent représenter, par des règles élémentaires, l'évolution d'une population de bactéries par exemple. Supposons partir d'une seule bactérie au temps 0. Au temps 1 cette bactérie va se dupliquer en plusieurs autres bactéries, identiques à la première, selon une loi aléatoire μ . Ensuite, au temps 2, les bactéries issues de la première vont se dupliquer de la même manière et indépendamment des autres, toujours en suivant la loi μ . Ainsi de suite, à chaque étape i , les bactéries de la i -ième génération vont se dupliquer indépendamment selon la même loi μ .

Pour donner une définition formelle nous devons d'abord donner l'ensemble dans lequel évolue un arbre de Bienaymé. Cet ensemble est celui des arbres plans, enracinés et localement finis. Un arbre est dit être *localement fini* si tous ces sommets ont un degré fini. À noter qu'un arbre localement fini peut avoir un nombre infini (mais toujours dénombrable) de sommets.

Définition I.C.1 : Arbre de Bienaymé

Soit μ une loi de probabilité à support dans \mathbb{N} . Un *arbre de Bienaymé* T de loi de reproduction μ est un arbre aléatoire vérifiant, pour tout arbre plan, enraciné et localement fini t :

$$\mathbb{P}(T = t) = \prod_{v \in t} \mu(k_v)$$

où le produit est sur l'ensemble des sommets de t et k_v désigne le nombre d'enfants de v dans t .

Remarque

Les arbres de Bienaymé sont aussi connus sous le nom d'arbres de Galton-Watson ou encore arbres de Bienaymé-Galton-Watson. La raison est que Galton et Watson, deux scientifiques britanniques, ont étudié ce processus d'évolution et ont publié un article [55] en 1874 énonçant le théorème d'extinction (Théorème I.C.2). Pendant un certain temps, Galton et Watson furent considérés, chez les probabilistes, comme les premiers à avoir formalisé et analysé un tel modèle. Mais en réalité, il s'avère qu'un mathématicien français, du nom de Bienaymé, avait déjà étudié un modèle similaire dès 1845 [23]. De plus, Galton et Watson ont commis une erreur dans [55] en disant que la probabilité d'extinction vaut toujours 1. Tandis que Bienaymé n'en a commis aucune. Ainsi nous suivons l'exemple de [2] et décidons de donner exclusivement le nom de Bienaymé à ces arbres.

Les deux scientifiques Galton et Watson, dans leur article de 1874 [55], voulaient répondre à un problème posé par Galton 1 an plus tôt :

A large nation, of whom we will only concern ourselves with adult males, N in number, and who each bear separate surnames colonise a district. Their law of population is such that, in each ge-



neration, a_0 per cent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five. Find (1) what proportion of their surnames will have become extinct after r generations; and (2) how many instances there will be of the surname being held by m persons.

Avec une terminologie plus moderne, le problème de Galton consiste donc à trouver la probabilité qu'un arbre de Bienaymé s'éteigne, c'est-à-dire, qu'il n'ait qu'un nombre fini de sommets. La réponse à ce problème est présentée dans le théorème suivant, appelé théorème d'extinction.

Théorème I.C.2 : Théorème d'extinction



Soit T un arbre de Bienaymé de loi de reproduction μ .

- Dans le régime surcritique, c'est-à-dire, si $\mathbb{E}[\mu] > 1$, alors T est infini avec une probabilité strictement positive.
- Dans le régime sous-critique, c'est-à-dire, si $\mathbb{E}[\mu] < 1$, alors T est fini presque sûrement.
- Dans le régime critique non-trivial, c'est-à-dire, si $\mathbb{E}[\mu] = 1$ et $\mu \neq \delta_1$, alors T est fini presque sûrement.

L'étude d'un arbre de Bienaymé peut se transposer à l'étude d'une marche aléatoire grâce au codage de Łukasiewicz. En effet, la marche de Łukasiewicz d'un arbre de Bienaymé correspond à une marche aléatoire arrêtée au premier temps d'atteinte de -1 . De plus conditionner l'arbre de Bienaymé à avoir n sommets revient à conditionner le temps d'atteinte de -1 à valoir n (voir par exemple [78, Corollaire 1.6]).

Théorème I.C.3 : Marche de Łukasiewicz d'un arbre de Bienaymé



Soit T un arbre de Bienaymé de loi de reproduction μ . Soit $(S_k)_{k \geq 0}$ une marche aléatoire démarrant en 0 ($S_0 = 0$) et de pas $X_i := S_i - S_{i-1}$ i.i.d de loi $\mathbb{P}(X_i = x) = \mu(x + 1)$ pour tout entier $x \geq -1$. Soit τ le temps d'atteinte de -1 , c'est-à-dire, $\tau := \min\{k \in \mathbb{N} : S_k = -1\}$ (avec la convention que $\min \emptyset = +\infty$). Alors

- La marche de Łukasiewicz de T a la même loi que $(S_k)_{0 \leq k < \tau+1}$.
- La marche de Łukasiewicz de T conditionné à avoir n sommets a la même loi que $(S_k)_{0 \leq k \leq n}$ conditionnée à ce que $\tau = n$.

Il existe deux résultats importants qui disent que les arbres de Cayley uniformes et les arbres plans enracinés uniformes sont des cas particuliers d'arbre de Bienaymé conditionnés à avoir une taille fixée.

Théorème I.C.4 : Lien avec arbres plans uniformes



Soit T un arbre de Bienaymé de loi de reproduction géométrique de paramètre $1/2$, $\mu \sim \mathcal{G}(1/2)$. Alors T conditionné à avoir n sommets a la même loi qu'un arbre plan enraciné aléatoire à n sommets de loi uniforme.

Théorème I.C.5 : Lien avec arbres de Cayley uniformes

Soit T un arbre de Bienaymé de loi de reproduction Poisson de paramètre 1, $\mu \sim \mathcal{P}(1)$. Si l'on conditionne T à avoir n sommets, que l'on numérote aléatoirement uniformément avec les entiers de 1 à n et que l'on oublie l'ordre planaire, alors T a la même loi qu'un arbre de Cayley aléatoire à n sommets de loi uniforme.

À noter que dans le cas d'une loi de reproduction géométrique de paramètre $1/2$ ou Poisson de paramètre 1, la moyenne de la loi vaut 1. Cela signifie que les arbres de Cayley uniformes et plans uniformes se situent dans le régime critique. En réalité, le Théorème I.C.4 resterait vrai en prenant une loi géométrique de paramètre $p \in]0, 1[$ quelconque. Le caractère critique serait perdu bien que la loi de l'arbre resterait inchangée. Nous discuterons, dans la Section C.iii, de la classification des lois de reproduction aboutissant aux mêmes arbres de Bienaymé conditionnés par la taille.

C.ii Application aux tuples de stationnement (suite)

Maintenant que la notion d'arbre de Bienaymé est traitée, nous pouvons continuer l'étude des tuples de stationnement initiée Section B.iii. Rappelons que l'on a défini une bijection entre les n -uplets de stationnement et les arbres de Cayley à $n+1$ sommets. Cette bijection additionnée au codage des arbres par la marche de Łukasiewicz (version parcours en largeur), permet de transposer l'étude d'un tuple de stationnement aléatoire uniforme à celle d'une marche aléatoire conditionnée. Grâce au Théorème I.C.5 et au Théorème I.C.3, la loi de cette marche aléatoire est explicite. Plus précisément, soit $(\pi_n(i))_{1 \leq i \leq n}$ un n -uplet de stationnement choisi uniformément au hasard dans \mathcal{S}_n . Soit $(X_i)_{i \geq 1}$ des variables aléatoires *i.i.d* de loi Poisson de paramètre 1. Pour tout $n \geq 0$ on pose $S_n := \sum_{i=1}^n (X_i - 1)$ et $\tau := \min\{n \geq 1 : S_n = -1\}$ le premier temps d'atteinte de -1 . Considérons la mesure de probabilité $\mathbb{P}_n := \mathbb{P}(\cdot | \tau = n+1)$ et posons $\mathbb{E}_n := \mathbb{E}[\cdot | \tau = n+1]$. Le Théorème I.B.5 et les propriétés de la marche de Łukasiewicz permettent alors de déduire le résultat suivant.

Théorème I.C.6 : B. [17]

Fixons $1 \leq k \leq n$ et $1 \leq i_1, \dots, i_k \leq n$. Soit $j_1 < \dots < j_r$ tels que $\{i_1, \dots, i_k\} = \{j_1, \dots, j_r\}$ (ainsi (j_1, \dots, j_r) correspond au vecteur (i_1, \dots, i_k) où les répétitions ont été enlevées, et r correspond au nombre d'éléments distincts de (i_1, \dots, i_k)). Soit $m_s = \#\{u : i_u = j_s\}$ Pour tout $1 \leq s \leq r$. Alors

$$\mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) = \frac{(n-k)!}{n!} \mathbb{E}_n \left[\prod_{s=1}^r (X_{j_s})_{m_s} \right] \quad (\text{C.1})$$

où $(x)_m := x(x-1) \cdots (x-m+1)$ désigne le symbole de Pochhammer.

Tuples de stationnement et loi uniforme. L'étude des propriétés du n -uplet de stationnement $(\pi_n(i))_{1 \leq i \leq n}$ choisi uniformément au hasard est récent et commence par l'article de Diaconis et Hicks [45] « Probabilizing parking functions » en 2017. Le titre de l'article est judicieux car il doit s'agir du premier article apportant une vision probabiliste aux tuples de stationnement. Plusieurs questions sont posées et résolues dans cet article comme : quelle est la loi du nombre de voitures voulant se garer à la première place ? quelle est la loi du nombre de i tel que $\pi_n(i) = \pi_n(i+1)$? Ou encore : quel est le comportement asymptotique des quantités



$\#\{k : \pi_n(k) \leq i\}$ et $\sum_{i=1}^n \pi_n(i)$? Les arguments employés dans [45] sont variés passant des fonctions génératrices à des éléments plus probabilistes. La question qui nous occupera sera celle du Corollaire 6 de [45]. Il y est montré que $\pi_n(1)$ est asymptotiquement semblable à une variable de loi uniforme sur $\llbracket 1, n \rrbracket$. Autrement dit, il semblerait que, pour n grand, la première voiture cherche à se garer indifféremment sur l'une des places disponibles. Diaconis et Hicks conjecturent que ce phénomène devrait rester valable pour les k premières voitures avec k fixé, lorsque n tend vers $+\infty$. Autrement dit, on ne pourrait pas distinguer $(\pi_n(1), \dots, \pi_n(k))$ d'un k -uplet $(U_n(1), \dots, U_n(k))$ de variables i.i.d de loi uniforme sur $\llbracket 1, n \rrbracket$. L'intuition nous dit que les voitures vont avoir tendance à vouloir se garer vers la gauche afin de garantir que tout le monde puisse se garer. Ainsi il est clair que ce phénomène ne peut perdurer pour de grandes valeurs de k . En effet pour $k \sim cn$ avec $c > 0$, les auteurs de [45] remarquent que ce phénomène de décalages vers la gauche est macroscopique. Durant cette thèse nous avons montré que, non seulement la conjecture de [45] est vraie, mais que le phénomène reste avéré pour $k = o(\sqrt{n})$. Formalisons cette discussion. Pour tout $k \in \llbracket 1, n \rrbracket$ notons

$$d_{TV}(k, n) := \frac{1}{2} \sum_{i_1, \dots, i_k=1}^n \left| \mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) - \frac{1}{n^k} \right|$$

la distance en variations totales entre $(\pi_n(1), \dots, \pi_n(k))$ et $(U_n(1), \dots, U_n(k))$. Diaconis et Hicks [45, Corollary 6] ont montré que $d_{TV}(1, n)$ tend vers 0 lorsque n tend vers l'infini et ont conjecturé que pour k fixé, $d_{TV}(k, n)$ devrait être un $O(k/\sqrt{n})$. Dans ce même article les auteurs étudient aussi la distance de Kolmogorov

$$d_K(k, n) := \max_{1 \leq i_1 \dots i_k \leq n} \left| \mathbb{P}(\pi_n(1) \leq i_1, \dots, \pi_n(k) \leq i_k) - \frac{i_1 \dots i_k}{n^k} \right|$$

et montrent que [45, Théorème 3] pour $k \in \llbracket 1, n \rrbracket$:

$$d_K(k, n) = O\left(k\sqrt{\frac{\log n}{n}} + \frac{k^2}{n}\right).$$

Comme dit précédemment, nous montrons une version plus forte de la conjecture de Diaconis et Hicks avec le théorème suivant.

Théorème I.C.7 : B. [17]		✖
(i) Si $k_n = o(\sqrt{n})$ alors	$d_{TV}(k_n, n) = O\left(\frac{k_n}{\sqrt{n}}\right).$	(C.2)
(ii) Si $k_n = o(n)$ et $\sqrt{n} = o(k_n)$ alors	$d_K(k_n, n) = O\left(\frac{\sqrt{n}}{k_n} + \left(\frac{k_n}{n}\right)^{0.19}\right).$	(C.3)

Remarque

Dans le Théorème I.C.7 (ii), \sqrt{n} est un $o(k_n)$. Puisque la fonction $k \mapsto d_K(k, n)$ est croissante à n fixé, la distance $d_K(k_n, n)$ converge encore vers 0 du moment que $k_n = o(n)$.

Les arguments de Diaconis et Hicks pour le cas $k = 1$ sont assez combinatoires tandis que ceux du Théorème I.C.7 sont plutôt probabilistes. À l'inverse du Théorème I.A.11, les probabilités généralisent la combinatoire. Il serait intéressant de savoir si la distance en variations totales converge encore vers 0 pour $k_n = o(n)$ ou s'il

existe un seuil intermédiaire entre $o(\sqrt{n})$ et $o(n)$ à partir duquel la convergence ne tient plus. La convergence de la distance de Kolmogorov peut laisser penser, dans une première heuristique, que la bonne réponse devrait être $o(n)$. Dans [47] il est montré que (C.2) est optimal pour $k_n = 1$ dans le sens où

$$d_{TV}(1, n) = \Theta\left(\frac{1}{\sqrt{n}}\right).$$

Une deuxième heuristique serait de penser qu'avec une accumulation des erreurs on pourrait avoir

$$d_{TV}(k_n, n) = \Theta\left(\frac{k_n}{\sqrt{n}}\right)$$

ce qui impliquerait que le seuil de convergence de la distance en variations totales serait $o(\sqrt{n})$. Le régime $k_n = o(\sqrt{n})$ est requis dans (C.2) car c'est lorsque $k_n \sim c\sqrt{n}$ que des collisions commencent à apparaître de manière macroscopique entre les voitures. Le fait d'avoir peu de collisions permet de simplifier grandement la formule (C.1) en supposant que i_1, \dots, i_k sont distincts. Dans ce cas on a

$$\mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) = \frac{(n-k)!}{n!} \mathbb{E}_n \left[\prod_{u=1}^k X_{i_u} \right].$$

Le problème est donc de savoir si les collisions apparaissant dans le régime \sqrt{n} empêchent la convergence en variations totales.

L'idée de la preuve de I.C.7 est d'utiliser le Théorème I.C.6 afin de se ramener à l'étude d'une marche aléatoire conditionnée. Ensuite on contrôle les moments de cette marche uniformément en temps. La majoration de ces moments est assez délicate et nous pensons qu'elle présente un intérêt en soi, indépendamment du problème (voir Proposition III.C.4). On déduit ensuite (C.2) par un jeu de majorations. Le cas de la distance de Kolmogorov utilise un argument supplémentaire d'échangeabilité dû à Bobkov [25, Théorème 1.1]. Ce résultat tire profit du fait que $(\pi_n(i))_{1 \leq i \leq n}$ reste invariant par permutation ce qui permet d'approximer la fonction de répartition de $(\pi_n(i))_{1 \leq i \leq k}$ par l'espérance du produit des fonctions de répartition marginales empiriques.

C.iii Arbres simplement générés

Cette section traite des arbres simplement générés qui sont une généralisation des arbres de Bienaymé conditionnés à avoir une taille fixée. Tous les éléments présentés ici proviennent de l'article, très complet, de Janson [68].

Définition I.C.8 : Arbre simplement généré

Soit $\mathbf{w} := (w_i)_{i \geq 0}$ une famille de nombres réels positifs. Un *arbre simplement généré* à n sommets avec poids \mathbf{w} est un arbre aléatoire T tel que pour tout arbre plan enraciné et fini t :

$$\mathbb{P}(T = t) = \frac{1}{Z_n} \left(\prod_{v \in t} w_{k_v} \right) \mathbb{1}_{|t|=n}$$

où $|t|$ désigne le nombre de sommets de t , k_v désigne le nombre d'enfants du sommet v et Z_n est la constante de normalisation

$$Z_n := \sum_{|t|=n} \prod_{v \in t} w_{k_v}.$$



Lorsque \mathbf{w} forme une loi de probabilités, c'est-à-dire, quand la somme des poids vaut 1, alors l'arbre T n'est rien d'autre qu'un arbre de Bienaymé conditionné à avoir n sommets avec loi de reproduction \mathbf{w} . Le modèle des arbres simplement générés inclut donc celui des arbres de Bienaymé conditionnés à avoir un nombre fixé de sommets. Ce modèle est même strictement plus grand puisque, par exemple, la suite de poids $w_i = i!$ donne un arbre simplement généré qui n'est équivalent à aucun arbre de Bienaymé conditionné. Dans la suite nous allons classifier les arbres simplement générés afin de déterminer lesquels sont équivalents à un arbre de Bienaymé conditionné. Soit $\rho \in [0, +\infty]$ le rayon de convergence de

$$\phi(x) := \sum_{i \geq 0} w_i x^i.$$

Dans [68, Lemme 3.1] Janson montre que, si $\rho > 0$, alors la fonction

$$\psi(x) := \frac{x\phi'(x)}{\phi(x)}$$

est croissante sur $[0, \rho[$ et on peut définir $\nu := \lim_{x \rightarrow \rho} \psi(x) \in]0, +\infty]$. On distingue alors trois régimes.

- Régime 1 : $\rho > 0$ et $\nu \geq 1$. Dans ce cas il existe un unique $\tau \in [0, \rho]$ tel que $\tau < +\infty$ et $\psi(\tau) = 1$.
- Régime 2 : $\rho > 0$ et $0 < \nu < 1$. Dans ce cas $\rho < +\infty$ et on pose $\tau := \rho$.
- Régime 3 : $\rho = 0$.

Dans le régime 1 et 2 on peut définir une loi de probabilité

$$\mu(k) := \frac{\tau^k w_k}{\phi(\tau)}. \tag{C.4}$$

Un résultat important de [68] est que, T , dans les régimes 1 et 2, a la même loi qu'un arbre de Bienaymé de loi de reproduction μ conditionné à avoir n sommets. Dans le régime 3, T n'a pas la loi d'un arbre de Bienaymé conditionné. À noter qu'un arbre de Bienaymé critique ou surcritique, conditionné à avoir n sommets, se trouve toujours dans le régime 1. Cela montre que, lorsque l'on étudie des arbres de Bienaymé conditionnés à avoir un nombre de sommets fixés, le cas surcritique n'apporte rien de plus et l'on peut se contenter des cas sous-critiques et critiques. Un arbre de Bienaymé sous-critique conditionné peut se trouver aussi bien dans le régime 1 que dans le 2.

Nous terminons cette section par une remarque technique. Pour que la Définition I.C.8 soit valide il faut que la quantité Z_n soit non nulle. Tout d'abord, supposons que $w_0 > 0$ et $w_k > 0$ pour un certain $k \geq 1$ sinon $Z_n = 0$ pour tout $n \geq 1$. Soit $\delta(\mathbf{w})$ le plus grand diviseur commun de $\{i \geq 0 : w_i > 0\}$. Le résultat suivant caractérise les n tels que $Z_n > 0$:

Lemme I.C.9 : Janson [68]



Si $Z_n > 0$ alors $n \equiv 1$ modulo $\delta(\mathbf{w})$. Inversement, il existe n_0 tel que pour tout $n \geq n_0$ vérifiant $n \equiv 1$ modulo $\delta(\mathbf{w})$, $Z_n > 0$.

Dans toute la suite nous supposerons que $w_0 > 0$, $w_k > 0$ pour un certain $k \geq 1$ et que n satisfait $n \geq n_0$ et $n \equiv 1 \pmod{\delta(\mathbf{w})}$. En particulier, nous supposerons tacitement ces conditions vérifiées lorsque nous conditionnerons un arbre de Bienaymé à avoir exactement n sommets.

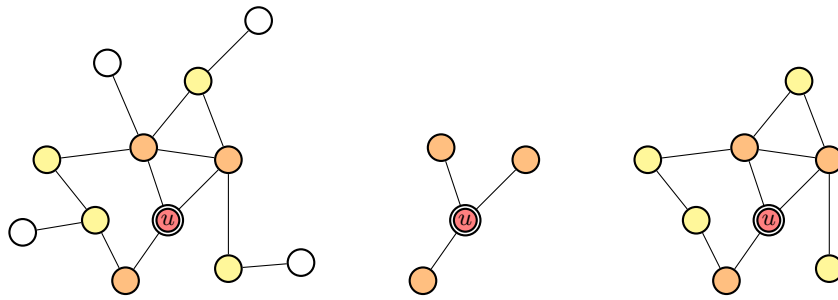


FIGURE D.1 – Représentation d'un graphe enraciné (G, u) à gauche et de ses voisinages $[(G, u)]_1$ au milieu et $[(G, u)]_2$ à droite. Les sommets à distance 0 de la racine (il n'y a que la racine elle-même u) sont en rouge, ceux à distance 1 sont en orange, ceux à distance 2 en jaune et les autres en blanc.

D Convergence de grands arbres

D.i Limites locales

L'une des questions les plus importantes dans l'étude des arbres aléatoires est celle de leur convergence lorsque leur taille croît vers l'infini. La notion de taille d'un arbre peut signifier plusieurs choses : le nombre de sommets, la hauteur de l'arbre, le nombre de feuilles, etc. Ici nous nous intéresserons uniquement au nombre de sommets. Autrement dit, dans ce document, un arbre de *taille* n sera un arbre avec n sommets. Tout comme il existe plusieurs notions de taille, il existe aussi plusieurs notions de convergence. Les deux notions principales sont celles de convergence locale, que nous développons dans cette section, et celle de convergence globale, que nous aborderons, brièvement, Section D.iii. La notion de convergence locale peut s'appliquer dans le cadre général des graphes et pas seulement des arbres. Soit (G, u) un graphe enraciné en u et $h \in \mathbb{N}$. On note $[(G, u)]_h$ le voisinage de u de taille h dans G , il s'agit du sous-graphe de G enraciné en u obtenu en tronquant les sommets et arêtes à distance au moins h de u . Plus formellement, les sommets de $[(G, u)]_h$ sont les sommets v de G tels que $d(u, v) \leq h$ et les arêtes de $[(G, u)]_h$ sont celles de G telles que les deux extrémités sont dans $[(G, u)]_h$ (voir Figure D.1). On dira qu'une suite de graphes enracinés $(G_n, u_n)_n$ converge vers un graphe enraciné (G, u) si tout voisinage de u_n dans G_n est égal au voisinage de u dans G pour tout n assez grand (voir Figure D.2). La convergence n'étant exigée qu'au voisinage des racines, elle ne rend compte que du comportement local de ces dernières et ne permet pas d'appréhender le comportement global des graphes. Une convergence locale n'informe donc pas sur le diamètre des graphes mais informe sur le degré des racines par exemple.

Définition I.D.1 : Convergence locale

Soit $((G_n, u_n))_{n \geq 1}$ une suite de graphes enracinés et (G, u) un graphe enraciné. On dit que la suite $((G_n, u_n))_{n \geq 1}$ converge localement vers (G, u) si et seulement si pour tout $h \in \mathbb{N}$, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $[(G_n, u_n)]_h$ et $[(G, u)]_h$ sont identiques (à un isomorphisme près).

Mentionnons un résultat important de convergence locale en ce qui concerne les arbres de Bienaymé. Rappelons que par définition, les arbres de Bienaymé sont considérés comme enracinés, la convergence locale sera donc sous-entendue être au voisinage de cette racine. Considérons un arbre de Bienaymé T_n , de loi de reproduction μ non dégénérée, conditionné à avoir n sommets. Supposons que l'on soit dans le cas critique,

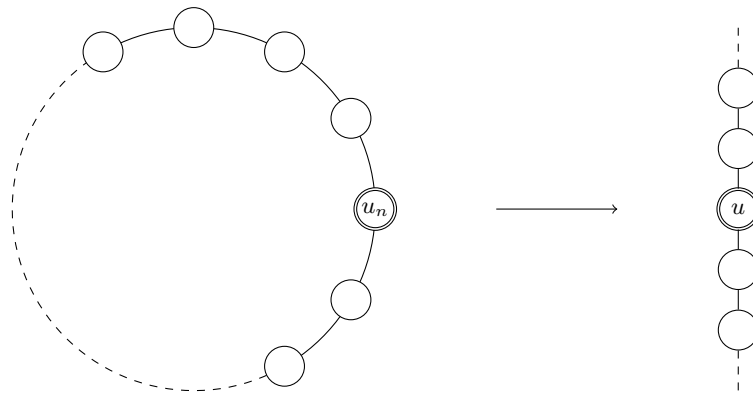


FIGURE D.2 – Exemple de convergence locale. Ici le graphe enraciné (G_n, u_n) correspond à un cercle formé par n sommets. La suite $(G_n, u_n)_n$ converge localement vers le graphe enraciné (G, u) formant une ligne infinie de sommets. La limite locale ne rend pas compte de la structure circulaire globale des graphes (G_n, u_n) .

c'est-à-dire que μ est de moyenne 1. Par le Théorème I.C.2, un arbre de Bienaymé de loi μ s'éteint presque sûrement. Ainsi conditionner cet arbre à avoir de nombreux sommets revient, en quelque sorte, à le conditionner à survivre longtemps. Ce conditionnement se concrétise en forçant un **unique** individu à chaque génération à avoir au moins un enfant tout en laissant les autres individus se reproduire normalement selon la loi μ . Cet individu spécial sur qui repose la survie de l'arbre est appelé un *mutant* et se reproduit selon la loi $\hat{\mu}$ qui correspond à la loi μ biaisée par la taille :

$$\hat{\mu}(k) := k\mu(k) \quad \forall k \in \mathbb{N}.$$

Le mutant de la $(i + 1)$ -génération est choisi uniformément au hasard parmi les enfants du mutant à la i -ième génération. Les autres individus, non mutants, se reproduisent selon la loi μ . Toujours par le Théorème I.C.2, il est presque sûr que les individus normaux auront une descendance finie. La survie de l'arbre est donc garantie uniquement grâce à la lignée des mutants qui forment une épine dorsale infinie dans l'arbre. L'arbre obtenu de la manière décrite ci-dessus est appelé *l'arbre de Kesten*. On obtient le résultat suivant (voir [1, Corollaire 3.20])

Théorème I.D.2 : Convergence vers l'arbre de Kesten



Pour tout entier $n \geq 1$, soit T_n un arbre de Bienaymé de loi de reproduction μ conditionné à avoir n sommets. Soit T^* l'arbre de Kesten défini plus haut. Si l'on suppose que μ est de moyenne 1 (cas critique), alors la convergence

$$T_n \xrightarrow[n \rightarrow \infty]{(d)} T^*$$

a lieu en loi selon la topologie de la convergence locale.

Kesten a montré [74] que le Théorème I.D.2 est valide dans le cas critique et sous-critique lorsque l'on conditionne la hauteur de l'arbre à être plus grande que n (plutôt que de conditionner le nombre de sommets à valoir n). Ce théorème de Kesten est le premier théorème de convergence locale d'arbres de Bienaymé conditionnés et c'est pourquoi l'arbre limite T^* porte son nom.

Remarque

À noter qu'avec notre définition de convergence locale, seule la forme des arbres de Bienaymé est considérée car l'ordre planaire est ignoré dans la définition de voisinage. La convergence de ce théorème resterait cependant correcte si l'on conservait l'ordre planaire mais cela nécessiterait alors un formalisme un peu plus lourd (par exemple le formalisme de Neveu [86]). Il le serait d'autant plus dans ce qui suit lorsque l'on ré-enracinera les arbres. Pour des raisons de simplicité nous ignorons l'ordre planaire ici. De plus, l'application principale qui nous intéressera, à savoir, le problème du plus grand ensemble d'indépendants, ne prend en compte que la forme des graphes. Cependant le lecteur pourra trouver tout le formalisme ainsi que les résultats les plus généraux de Stufler [97] dans le Chapitre IV en ce qui concerne la convergence locale d'arbre ré-enracinés.

Nous poursuivons cette section en nous posant la question suivante : que se passe-t-il si l'on ré-enracine un arbre simplement généré uniformément au hasard ? Le résultat du Théorème I.D.2 ne nous aidera pas à répondre puisqu'a priori un sommet de T_n choisi uniformément au hasard a toutes les chances d'être très éloigné de la racine initiale de T_n et donc le caractère local de la convergence nous restreint. Pour répondre à la question nous nous appuyons sur les résultats de Stufler [97]. Soit \mathbf{w} une suite de poids et T_n un arbre simplement généré de taille n avec poids \mathbf{w} . Nous reprenons les notations de la Section C.iii. Soit U_n un sommet aléatoire qui, conditionnellement à T_n , est choisi uniformément au hasard dans T_n . Stufler montre que, sous certaines conditions, l'arbre T_n ré-enraciné en U_n converge localement. La limite locale dépend du régime dans lequel \mathbf{w} se trouve. Décrivons d'abord les 3 limites différentes T_1^* , T_2^* and T_3^* correspondant respectivement aux régimes 1, 2 et 3 et voyons ensuite les conditions de convergence (voir Figure D.3).

Rappelons que, dans le régime 1 et 2, μ est une loi de probabilité définie par (C.4). Soit T un arbre de Bienaymé de loi μ . Soit $\hat{\mu}$ la loi de probabilité définie sur $\mathbb{N} \cup \{\infty\}$ par

$$\hat{\mu}_k = k\mu_k \quad \forall k \in \mathbb{N} \quad \text{et} \quad \hat{\mu}_\infty = 1 - m$$

où $m := \min(1, \nu)$ est la moyenne de μ .

- Commençons par définir T_1^* dans le premier régime. Remarquons que dans ce cas $\hat{\mu}_\infty = 0$. On commence par un sommet racine u_0 auquel on attache une copie indépendante de T . On attache ensuite à u_0 une chaîne infinie de sommets alignés u_1, u_2 , etc. Pour tout $k \geq 1$, u_k se reproduit selon $\hat{\mu}$. Enfin, on attache une copie indépendante de T à tous les enfants de u_k , hormis u_{k-1} .
- Continuons avec T_2^* dans le deuxième régime. Dans ce cas $\hat{\mu}_\infty > 0$. On commence par un sommet racine u_0 auquel on attache une copie indépendante de T . On attache ensuite à u_0 une chaîne infinie de sommets alignés u_1, u_2 , etc. Pour tout $k \geq 1$, u_k se reproduit selon $\hat{\mu}$. Presque sûrement, il existe deux entiers $1 \leq i < j$ tels que $u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_{j-1}$ sont de degré fini et u_i et u_j sont de degré infini. Ensuite, on attache une copie indépendante de T à tous les enfants de u_k , hormis u_{k-1} . Enfin, l'arbre T_2^* est le sous-arbre obtenu en gardant uniquement les descendants de u_{j-1} .
- Terminons avec T_3^* dans le troisième régime. Il s'agit simplement d'une racine u_0 à laquelle on attache un sommet u_1 possédant une infinité de voisins qui sont tous des feuilles.

Maintenant définissons la condition technique sous laquelle la convergence locale a lieu. Cette condition



est appelée *condensation complète* et correspond à

$$\Delta(T_n) = (1 - m)n + n\mathcal{E}_n$$

où $\Delta(T_n)$ est le degré maximal d'un sommet de T_n et \mathcal{E}_n est une variable aléatoire convergeant vers 0 en probabilité. Par exemple, la condensation complète se produit dans le régime 2 si μ est de la forme $\mu(k) = \ell(k)k^{-(1+\theta)}$ avec $\theta > 1$ et ℓ une fonction à variations lentes (voir [76]). La condensation complète se produit aussi dans le régime 3 si $w_k = k!^\alpha$ avec $\alpha > 0$ (voir [67, Ex. 19.36]).

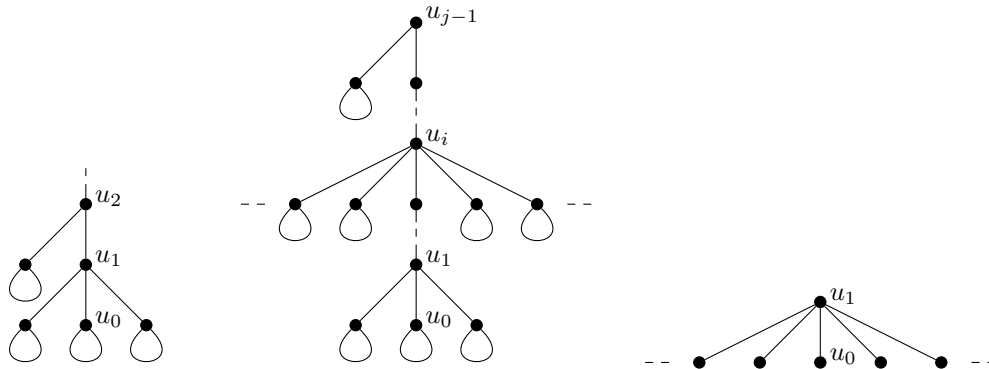


FIGURE D.3 – Représentation des arbres T_1^* , T_2^* and T_3^* , respectivement, de gauche à droite. Chaque boucle représente une copie indépendante d'un arbre de Bienaymé de loi μ .

Théorème I.D.3 : Stufler [97]



Supposons que l'on soit dans le régime $i = 1$ ou dans le régime $i \in \{2, 3\}$ avec condensation complète. Alors la convergence

$$(T_n, U_n) \xrightarrow[n \rightarrow \infty]{\text{loi}} T_i^*$$

a lieu en loi selon la topologie de la convergence locale.

D.ii Application au problème du plus grand ensemble d'indépendants

Un problème connu en théorie des graphes est le problème du plus grand ensemble d'indépendants. Ce problème consiste à trouver, dans un graphe donné, le plus grand sous-ensemble de sommets tel qu'aucune paire de sommets n'est reliée par une arête. Plus formellement,

Définition I.D.4 : (Plus grand) ensemble d'indépendants

Soit $G = (S, A)$ un graphe fini et $I \subset S$. On dit que I est un

- ensemble d'indépendants si pour tout $u, v \in I$, $\{u, v\} \notin A$.
- plus grand ensemble d'indépendants (PGEI) si c'est un ensemble d'indépendants de cardinal maximal.

On note $\alpha(G)$ la taille d'un PGEI.

Remarque

Noter que l'on dit toujours **un** PGEI car a priori il pourrait y avoir plusieurs PGEI dans un même graphe. Noter aussi que l'on considère ici la notion de maximalité au sens du cardinal et non au sens de l'inclusion ensembliste. Les ensembles d'indépendants maximaux selon le second sens ont aussi été étudiés (voir par exemple [80]) mais ici nous nous focaliserons uniquement sur le premier sens. D'ailleurs pour éviter toute confusion nous proposons d'utiliser l'expression « plus grand ensemble d'indépendants » pour le cardinal et « ensemble d'indépendants maximal » pour l'inclusion, ce qui respecte la nature partielle de l'ordre induit par l'inclusion.

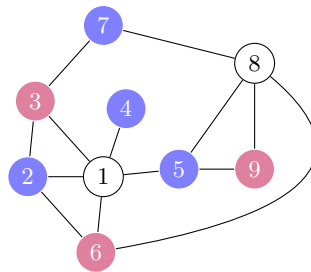


FIGURE D.4 – Exemple d'un graphe à 9 sommets. Les sommets violets forment un ensemble d'indépendants puisque 3, 6 et 9 ne sont pas reliés. Cependant ils ne forment pas un PGEI car les sommets bleus sont plus nombreux et forment aussi un ensemble d'indépendants. Ces derniers forment même un PGEI : en effet il n'existe pas d'ensemble d'indépendants contenant 5 ou plus de sommets dans ce graphe. Ainsi $\alpha(G) = 4$.

Motivations et bibliographie Le problème du PGEI consiste à trouver un PGEI, ou a minima la taille d'un PGEI, dans une famille donnée de graphes. Ce problème est bien étudié en théorie de la complexité algorithmique. Il est connu que décider si $\alpha(G) \geq k$, pour un graphe quelconque G et un entier $k \in \mathbb{N}$, est NP-complet, c'est-à-dire, NP et NP-difficile (voir par exemple [56, Sec. 3.1.3]). Beaucoup d'efforts ont été mis en oeuvre afin de décrire des algorithmes efficaces calculant α pour des graphes quelconques [91, 102]. À notre connaissance, le meilleur algorithme actuel est celui de Xiao et Nagamochi [102] en $O(1,1996^n n^c)$ où c est une constante. Dans des classes plus restreintes de graphes, les algorithmes peuvent se spécialiser et devenir plus performants réduisant ainsi la complexité du calcul. C'est pourquoi le problème du PGEI a été étudié dans des classes spécifiques de graphes : par exemple les arbres cubiques (c'est-à-dire des arbres dont le degré d'un sommet est au plus 3) [101], les graphes « sans pinces » [84] et les graphes P_3 -évitant [79]. Dans la communauté combinatoire et probabiliste ce problème a aussi été abordé. Par exemple, des méthodes probabilistes ont été utilisées pour déterminer des bornes supérieures sur les graphes cubiques [26]. Une borne inférieure générale, due à Caro [31] et Wei [99], peut se démontrer élégamment à partir d'un argument probabiliste simple, dû à Alon et Spencer [8], que nous présentons ici.

Lemme I.D.5 : Caro [31] et Wei [99]

Pour un graphe $G = (S, A)$ on a la borne inférieure suivante

$$\alpha(G) \geq \sum_{u \in S} \frac{1}{\deg(u) + 1}.$$

Preuve

Cette preuve probabiliste est due à Alon et Spencer [8]. Soit \leq un ordre total sur S choisi uniformément au hasard. Soit


$$I := \{u \in S : \forall v \in S, \{u, v\} \in A \implies u \leq v\}.$$

Alors

$$\mathbb{E}[\text{Card}(I)] = \mathbb{E}\left[\sum_{u \in S} \mathbb{1}_{u \in I}\right] = \sum_{u \in S} \mathbb{P}(u \in I).$$

Remarquons que $u \in I$ si et seulement si u est le plus petit, selon \leq , élément parmi ses voisins. Ainsi

$$\mathbb{P}(u \in I) = \frac{1}{\deg(u) + 1}.$$

On peut alors trouver un ordre total \leq tel que $\text{Card}(I) \geq \sum_{u \in S} \frac{1}{\deg(u) + 1}$. On conclut en voyant que I est un ensemble d'indépendants. 

Des résultats asymptotiques concernant le problème du PGEI ont été prouvés pour certaines classes d'arbres aléatoires, par exemple pour les arbres de Bienaymé conditionnés [44], pour les arbres simplement générés [13], les arbres récursifs et les arbres de recherche binaires [54] et aussi pour des classes plus larges d'arbres construits à partir d'un processus de branchement de type Crump-Mode-Jagers [70].

Problèmes reliés. Le problème du PGEI est relié à d'autres problèmes généraux de la théorie des graphes.

- Il est relié au problème du plus petit recouvrement qui consiste à trouver un ensemble de sommets, de cardinal minimal, tel que chaque arête du graphe est adjacente à, au moins, un sommet de l'ensemble. Dans un graphe G , un ensemble de sommet est un PGEI si et seulement si son complémentaire est un plus petit recouvrement. Ces deux problèmes sont donc clairement équivalents.
- Il est aussi lié au problème de la plus grande clique, qui est un problème réputé être NP-complet (voir par exemple [72]). Cela se voit facilement en considérant le graphe complémentaire $G^c := (S, A^c)$ de $G = (S, A)$ où A^c correspond au complémentaire de A (voir Figure D.5). En effet, G a un ensemble d'indépendants de cardinal k ou plus si et seulement si G^c a une clique de cardinal k ou plus. De cette manière on déduit que le problème du PGEI est aussi NP-complet.
- Un autre problème connexe est celui du plus grand appariement dans un graphe, c'est-à-dire, du plus grand ensemble d'arêtes que l'on puisse trouver qui ne partagent pas de sommets en commun. Le cardinal d'un plus grand appariement d'un graphe G à n sommets est toujours plus petit que $n - \alpha(G)$. De plus il est exactement égal à $n - \alpha(G)$ si G est biparti [77].



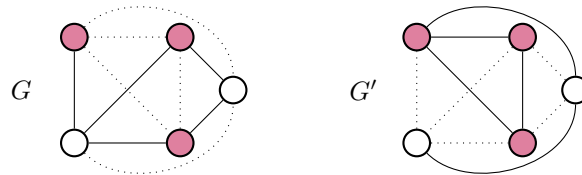


FIGURE D.5 – Exemple d'un graphe G à gauche et de son complémentaire G^c à droite. Les trois sommets violets forment un ensemble d'indépendants pour G et une clique pour G^c .

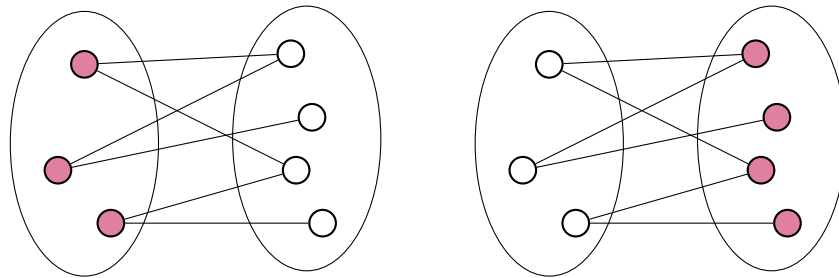


FIGURE D.6 – Exemple d'un graphe biparti. À gauche, les sommets de la première partie sont coloriés en rouge et forment un ensemble d'indépendants. À droite, les sommets de la deuxième partie sont coloriés en rouge et forment aussi un ensemble d'indépendants. Dans l'un des deux cas, le nombre de sommets rouges doit être supérieur aux nombre total de sommets divisé par 2.

Notre approche. Notre but est d'étudier le problème du PGEI dans un arbre. La structure d'arbre permet de simplifier considérablement l'étude du problème étant donné l'absence de cycles. Par exemple une borne inférieure très simple sur α consiste à dire que pour un arbres T à n sommets, on a $\alpha(T) \geq n/2$. En effet, les sommets à hauteur paire, ainsi que ceux à hauteur impaire, de T forment un ensemble d'indépendants. Puisque l'un des deux ensembles contient au moins $n/2$ éléments on en déduit la borne. Plus généralement, cet argument s'applique à un graphe biparti G à n noeuds : $\alpha(G) \geq n/2$ (voir Figure D.6). Plus précisément nous souhaitons obtenir des résultats asymptotiques sur les arbres simplement générés introduits Section C.iii. Comme dit plus haut, les auteurs de [13] obtiennent déjà des résultats dans ce cas.

Théorème I.D.6 : Banderier, Kuba & Panholzer [13]



Soit T_n un arbre simplement généré à n sommets sous le régime 1. Alors

$$\mathbb{E}[\alpha(T_n)] = nq + O(1) \quad \text{et} \quad V(\alpha(T_n)) = nv + O(1)$$

où q et v sont des constantes explicites.



Notre approche est très différente de celle de [13] car passe par l'étude d'un tricoloriage des arbres due à Zito [105]. Plus précisément, un *tricoloriage* d'un arbre fini t consiste à colorier les sommets de t en trois couleurs vert, rouge et orange de la manière suivante. Pour tout sommet v de t , on colorie v en

- vert si v n'appartient à aucun PGEI de t ,
- rouge si v appartient à tous les PGEI de t et
- orange sinon.

On note $n_v(t)$, $n_r(t)$ et $n_o(t)$, respectivement, le nombre de sommets verts, rouges et oranges de t . Il a été remarqué dans [40] que $\alpha(t)$ est égal à $n_r(t) + n_o(t)/2$. Dans [40], les auteurs utilisent ce fait pour calculer explicitement l'espérance du nombre de sommets verts, rouges et oranges dans des arbres de Cayley uniformes. À notre connaissance il n'existe aucun calcul explicite de ce genre pour d'autres classes d'arbres aléatoires. Ce calcul est permis grâce à une propriété bien spécifique des arbres de Cayley uniformes : ils sont invariants en loi par ré-enracinement uniforme. Expliquons brièvement en quoi cette propriété est pertinente. Supposons que l'on veuille étudier l'espérance du nombre de sommets verts dans un arbre aléatoire T_n à n sommets. Soit U_n un sommet choisi uniformément au hasard dans T_n . Alors on a

$$\mathbb{E}[n_v(T_n)] = \mathbb{E}\left[\sum_{v \in T_n} \mathbb{1}_{\{v \text{ est vert}\}}\right] = n\mathbb{P}(U_n \text{ est vert}).$$

Dans le cas où T_n est un arbre de Cayley à n sommets uniforme, on a alors que $\mathbb{P}(U_n \text{ est vert})$ est égal à $\mathbb{P}(\text{la racine est verte})$. Cette dernière quantité peut alors se calculer grâce à une construction récursive des arbres de Cayley : pour construire un arbre de Cayley à n sommets, on part d'une racine à laquelle on colle d'autres arbres de Cayley dont la somme des tailles vaut $n-1$. La construction récursive des arbres est d'ailleurs un ingrédient essentiel dans l'approche de [40], [13] ainsi que la nôtre. Nous avons vu avec le Théorème I.C.5 qu'un arbre de Cayley uniforme n'est qu'un cas particulier d'arbre de Bienaymé conditionné qui lui-même n'est qu'un cas particulier d'arbre simplement généré. Si nous voulions étudier le comportement de $n_v(T_n)$, $n_r(T_n)$ et $n_o(T_n)$ pour un arbre simplement généré T_n à n sommets, il faudrait comprendre le comportement de T_n après ré-enracinement uniforme. C'est là qu'interviennent les résultats de Stufler (cf Théorème I.D.3). Ces résultats donne l'existence, dans le régime $i = 1$ et dans les régimes $i = 2, 3$ sous condensation complète, d'un arbre limite T_i^* vers lequel converge localement T_n ré-enraciné en U_n . Cela nous permettra, après un certain travail, de montrer que $\mathbb{P}(U_n \text{ est vert})$ converge vers $\mathbb{P}(\text{la racine de } T_i^* \text{ est verte})$. Cette conclusion n'est pas triviale car, a priori, le tricoloriage d'un arbre dépend du comportement global de ce dernier, or les résultats de Stufler ne sont que des convergences locales. Il s'avère que cela sera suffisant car la couleur d'un sommet ne dépend en réalité que du comportement de son voisinage. Ce fait, a priori non évident, vient de la structure particulière de l'arbre limite T_i^* . Plus précisément, dans le régime $i = 1$, T_1^* possède une unique branche infinie et il est presque sûre qu'un sommet sur cette branche soit vert. Ainsi ce qui se passe au-dessus de ce sommet importe peu dans l'étude des sommets du dessous. Dans le régime $i = 2$ ou $i = 3$ avec condensation complète, il existe un sommet de degré infini, à distance finie de la racine de T_i^* , qui sera nécessairement vert. Encore une fois, cela implique que la couleur des sommets en dessous de ce dernier ne dépendent pas de ce qui se passe au dessus. Présentons enfin notre résultat principal concernant cette section.

Théorème I.D.7 : B. [15]



Soit T_n un arbre simplement généré avec n sommets de poids $\mathbf{w} = (w_i)_{i \geq 0}$. Rappelons que, dans le régime 1 et 2, G est la fonction génératrice de μ définie par (C.4). Soit q l'unique solution de $q = G(1 - q)$ dans $[0, 1]$.

1. Dans le régime 1 et dans le régime 2 avec condensation complète, les convergences suivantes sont valables dans L^p pour tout $p > 0$:

$$\frac{n_v(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{1 - q + (1 - 2q)G'(1 - q)}{1 + G'(1 - q)}, \quad \frac{n_o(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{2qG'(1 - q)}{1 + G'(1 - q)},$$

$$\frac{n_r(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{q}{1 + G'(1 - q)}.$$

2. dans le régime 3 avec condensation complète, les convergences suivantes sont valables dans L^p pour tout $p > 0$:

$$\frac{n_v(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 0, \quad \frac{n_o(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 0, \quad \frac{n_r(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 1.$$

Les termes limites dans le Théorème I.D.7 correspondent aux probabilités que la racine de T_i^* soit d'une certaine couleur. Afin de montrer le Théorème I.D.7 nous montrons d'abord la convergence de l'espérance du nombre de sommets d'une certaine couleur en suivant la méthodologie expliquée ci-dessus. Afin d'obtenir des convergence L^p et pas seulement des convergences de moment d'ordre 1, nous montrons aussi la convergence du moment d'ordre 2. Pour cela nous appliquons le même principe que pour le moment d'ordre 1 mais cette fois en prenant deux sommets i.i.d U_n et V_n uniformément au hasard dans T_n . Nous montrons ensuite, à l'aide de [68, Théorème. 11.7], que les couleurs de U_n et V_n sont asymptotiquement indépendantes, ainsi le moment d'ordre 2 est équivalent au moment d'ordre 1 au carré.

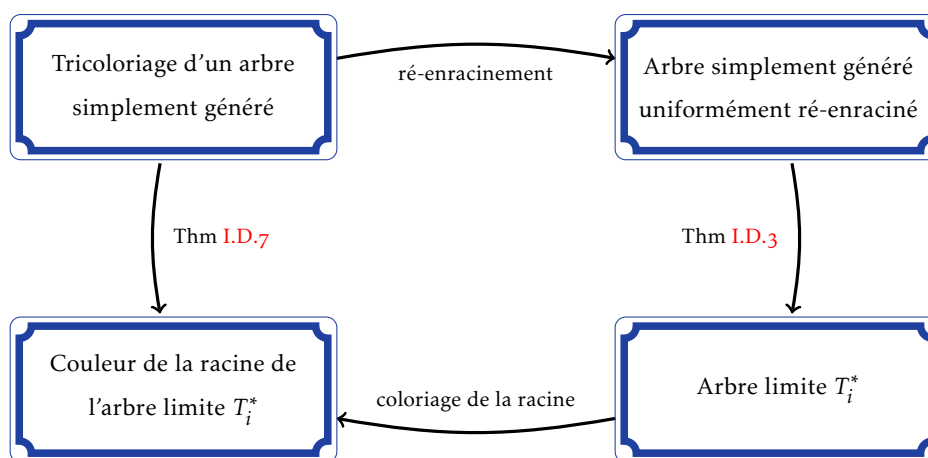


FIGURE D.7 – Stratégie de l'étude du tricoloriage d'un arbre de Bienaymé conditionné

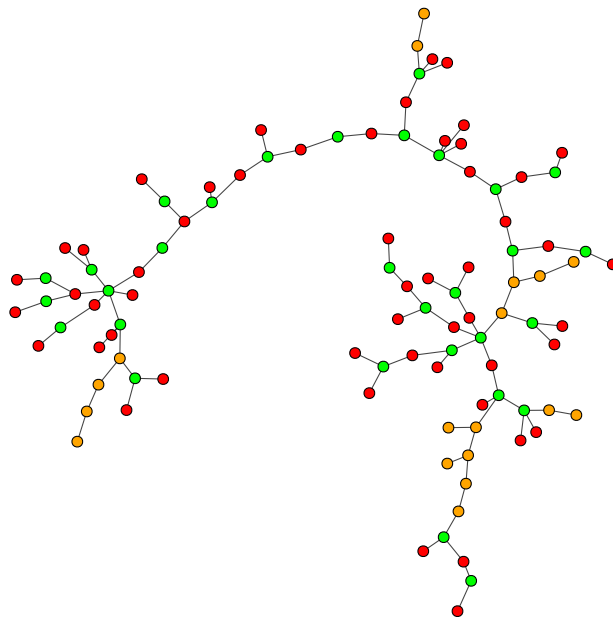


FIGURE D.8 – Tricoloriage d’un arbre de Cayley aléatoire de loi uniforme avec 100 sommets. L’algorithme de coloriage utilisé pour cette figure vient de Chapoton [33, Annexe A].

D.iii Limites d’échelles

Nous avons vu dans la Section D.i, des résultats de convergence locale des arbres de Bienaymé conditionnés. Nous voyons ici des résultats de convergence globale qui rendent compte de la structure entière de l’arbre et pas juste d’un voisinage autour d’un point. Ces résultats nous permettront d’étudier dans la section suivante le processus de lamination associé aux factorisations monotones du cycle. Les résultats de convergence globale des arbres de Bienaymé conditionnés font intervenir l’excursion brownienne ainsi que l’arbre brownien qui est défini à partir de cette dernière. Pour faire simple, l’excursion brownienne correspond au mouvement brownien (B_t) pris entre 0 et 1 conditionné par l’évènement $\{B_0 = B_1 = 0 \text{ et } B_t > 0 \forall t \in]0,1[\}$. L’excursion brownienne peut être obtenue en prenant un mouvement brownien entre le dernier instant où il touche 0 sur $[0,1]$ et le premier où il touche 0 sur $[1,+\infty[$. Plus formellement

Définition I.D.8 : Excursion brownienne

Soit $(B_t)_{t \geq 0}$ un mouvement brownien démarrant en 0. Soit

$$g := \max\{s \in [0,1] : B_s = 0\} \quad \text{et} \quad d := \min\{s \in [1,+\infty[: B_s = 0\}.$$

On définit alors l’excursion brownienne, notée $e = (e_t)_{0 \leq t \leq 1}$, par

$$e_t := \frac{B_{g+t(d-g)}}{\sqrt{d-g}} \quad \forall t \in [0,1].$$

Remarque

Une autre manière d'obtenir l'excursion brownienne est d'appliquer une transformée de Vervaat au pont brownien [81]. Le pont brownien correspond au mouvement brownien (B_t) conditionné par $\{B_0 = B_1 = 0\}$. La transformée de Vervaat consiste alors à prendre le premier instant t tel que le pont atteigne son minimum à l'instant t et ensuite de décaler le pont en le faisant démarrer à cet instant t . De cette manière on obtient une excursion de même loi que e .

L'excursion brownienne est un objet important car constitue la limite de la marche de Łukasiewicz d'un arbre de Bienaymé critique conditionné à variance finie. Plus précisément

Théorème I.D.9 : Convergence vers l'excursion brownienne

Soit $(S_k^n)_{0 \leq k \leq n}$ la marche de Łukasiewicz d'un arbre de Bienaymé conditionné à avoir n sommets de loi de reproduction μ critique. Supposons que μ à une variance finie notée σ^2 . Alors on a la convergence

$$\frac{(S_{tn}^n)_{0 \leq t \leq n}}{\sigma \sqrt{n}} \xrightarrow[n \rightarrow \infty]{} e.$$

en loi pour la topologie de la norme infinie.

En réalité ce théorème dit que le théorème bien connu de Donsker [46], qui stipule qu'une marche aléatoire centrée de variance finie converge, après renormalisation, vers un mouvement brownien, reste valable même après conditionnement par l'évènement « la marche reste positive sur $[0, n]$ et atteint 0 en n ». Il existe des généralisations du Théorème I.D.9 dans le cas de lois de reproduction de variance infinie et plus particulièrement dans le cas de lois de reproduction dans le domaine d'attraction d'une loi stable. Dans ce cas l'objet limite n'est plus l'excursion brownienne mais l'excursion d'un processus de Lévy. Cette généralisation a permis à Féray et Kortchemski [51] d'étudier le comportement asymptotique des laminations d'une factorisation minimale. Dans le cas des factorisations monotones nous n'aurons besoin que du cas de la variance finie.

La convergence du Théorème I.D.9 est la porte d'entrée menant à la convergence des arbres de Bienaymé conditionnés vers l'arbre brownien. Nous ne détaillons pas ces résultats puisqu'ils ne nous servent pas directement dans nos applications.

D.iv Application aux factorisation monotones du cycle (suite)

Le but de cette section est d'étudier le comportement asymptotique d'une factorisation minimale $(\tau_1^n, \dots, \tau_{n-1}^n)$, choisie au hasard dans \mathcal{F}_n où $\tau_i^n = (a_i^n, b_i^n)$ avec $a_i^n < b_i^n$. Plus précisément nous cherchons à étudier son processus de lamination déjà évoqué plusieurs fois.



Définition I.D.10 : Corde et lamination

Une *corde* est un segment de \mathbb{C} tel que les deux extrémités sont sur le cercle de centre 0 et de rayon 1. En somme, une corde est un ensemble de la forme

$$[[u, v]] := \{t \exp(-2i\pi u) + (1 - t) \exp(-2i\pi v) : t \in [0, 1]\}$$

Avec $u, v \in [0, 1]$. Deux cordes $[[u, v]]$ et $[[a, b]]$ sont dites *non croisées* si elles ne s'intersectent pas à l'intérieur du disque unité, autrement dit, si $[[u, v]] \cap [[a, b]] \subset \{u, v\}$. Enfin, une *lamination* est un sous-ensemble compact du disque unité qui peut s'écrire comme une union de cordes non croisées.

Par l'étude le comportement géométrique, nous entendons, l'étude du processus de lamination associé défini comme suit.

Définition I.D.11 : Processus discret de lamination

Pour $1 \leq k \leq n - 1$, on pose

$$L_k^n := \bigcup_{i \leq k} \left[\left[\frac{a_i^n}{n}, \frac{b_i^n}{n} \right] \right]$$

On pose aussi $L_0^n := [[0, 0]] = \{(1, 0)\}$ et $L_n^n := L_{n-1}^n$. Le processus $L^n := (L_{[nt]}^n)_{0 \leq t \leq 1}$ est appelé le *processus de lamination associé à $(\tau_1^n, \dots, \tau_{n-1}^n)$* ou, plus simplement, le *processus discret de lamination* (pour mettre en avant le fait que L^n est une union finie de cordes apparaissant à des temps discrets).

La dénomination de « lamination » est justifiée pour L^n car, d'après l'article [58], pour tout k les cordes de L_k^n sont non croisées.

Bibliographie et motivations Féray et Kortchemski [51] sont les premiers à avoir étudié le processus L^n dans le cas où $(\tau_1^n, \dots, \tau_{n-1}^n)$ est une factorisation minimales choisie uniformément au hasard dans \mathcal{F}_n . Dans ce cas ils obtiennent

Théorème I.D.12 : Féray & Kortchemski [51]



Pour tout $c > 0$ la convergence

$$L_{[c\sqrt{n}]}^n \xrightarrow{n \rightarrow \infty} \mathbb{L}_c$$

a lieu en loi pour la topologie de Hausdorff, où \mathbb{L}_c est un certaine lamination aléatoire.

Ce résultat fonctionne à c fixé. Une généralisation viendra ensuite de Thévenin [98] qui montre une convergence en processus, c'est-à-dire, en permettant la variation de c .

Théorème I.D.13 : Thévenin [98]

La convergence

$$(L^n_{\lfloor c\sqrt{n} \rfloor})_{c \geq 0} \xrightarrow{n \rightarrow \infty} (\mathbb{L}_c)_{c \geq 0}$$

a lieu en loi pour la topologie J_1 de Skorokhod.

Notre approche. Dans cette thèse nous avons cherché à obtenir un résultat similaire au Théorème I.D.13 lorsque $(\tau_n^1, \dots, \tau_n^n)$ est choisie uniformément au hasard dans \mathcal{F}_n^\downarrow (et non plus dans \mathcal{F}_n). Afin d'y parvenir nous exploitons la bijection entre factorisations croissantes et arbres plans décrite Théorème I.B.7. La marche de Łukasiewicz associée à un arbre plan uniforme, qui peut se voir comme un arbre de Bienaymé conditionné par le Théorème I.C.4, converge, après renormalisation, vers l'excursion brownienne d'après le Théorème I.D.9. La plus grande partie du travail consistera alors à montrer que le procédé qui permet de passer d'une excursion à la lamination est continue en \mathfrak{e} . Ce travail topologique aboutit au Théorème I.D.15.

Définition I.D.14 : Processus linéaire brownien de lamination

Le *processus linéaire brownien de lamination*, noté $(L_t(\mathfrak{e}))_{0 \leq t \leq 1}$, est un processus aléatoire à valeurs dans l'ensemble des laminations qui est obtenu de la manière suivante. Pour tout $t \in [0, 1]$, $L_t(\mathfrak{e})$ est l'union des cordes $[[u, v]]$ telles que $u, v \geq 1 - t$ et $\mathfrak{e}(u) = \mathfrak{e}(v) = \min_{[u, v]} \mathfrak{e}$.

Voir Figure D.9 pour illustrer la définition du processus linéaire brownien de lamination. Pour tout $t \in [0, 1]$, $L_t(\mathfrak{e})$ est bien une lamination car les cordes qui la composent ne se croisent pas dans l'intérieur du disque unité. Il est intéressant de noter que des triangles se dessinent aux endroits où \mathfrak{e} possède un minimum local. En effet si \mathfrak{e} atteint un minimum local en $u \in]0, 1[$ alors il existe $a < u < b$ tels que $[[a, u]]$ et $[[u, b]]$ seront deux cordes de $L_t(\mathfrak{e})$ pour $1 - t \leq a$. Les propriétés de l'excursion brownienne font que les minima locaux sont uniques, ce qui implique qu'aucun polygone, autre que le triangle, n'apparaîtra dans le processus linéaire brownien de lamination. La lamination aléatoire $L_1(\mathfrak{e})$ est appelée la *triangulation brownienne* et a été introduite par Aldous [5, 6].

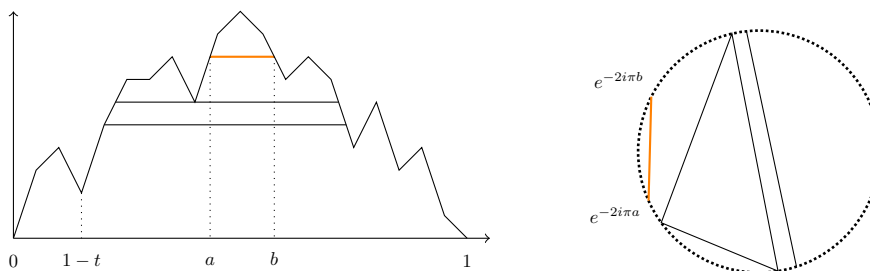


FIGURE D.9 – Illustration schématique de la construction du processus linéaire brownien de lamination. Ici sont dessinées trois cordes de $L_t(\mathfrak{e})$ dont une mise en valeur en orange.



Théorème I.D.15 : B. [16] ✖

La convergence

$$(L_{\lfloor nt \rfloor}^n)_{0 \leq t \leq 1} \xrightarrow{n \rightarrow \infty} (L_t(\epsilon))_{0 \leq t \leq 1}$$

a lieu en loi pour la topologie J_1 de Skorokhod, où $(L_t(\epsilon))_{0 \leq t \leq 1}$ est le processus linéaire brownien de lamination.

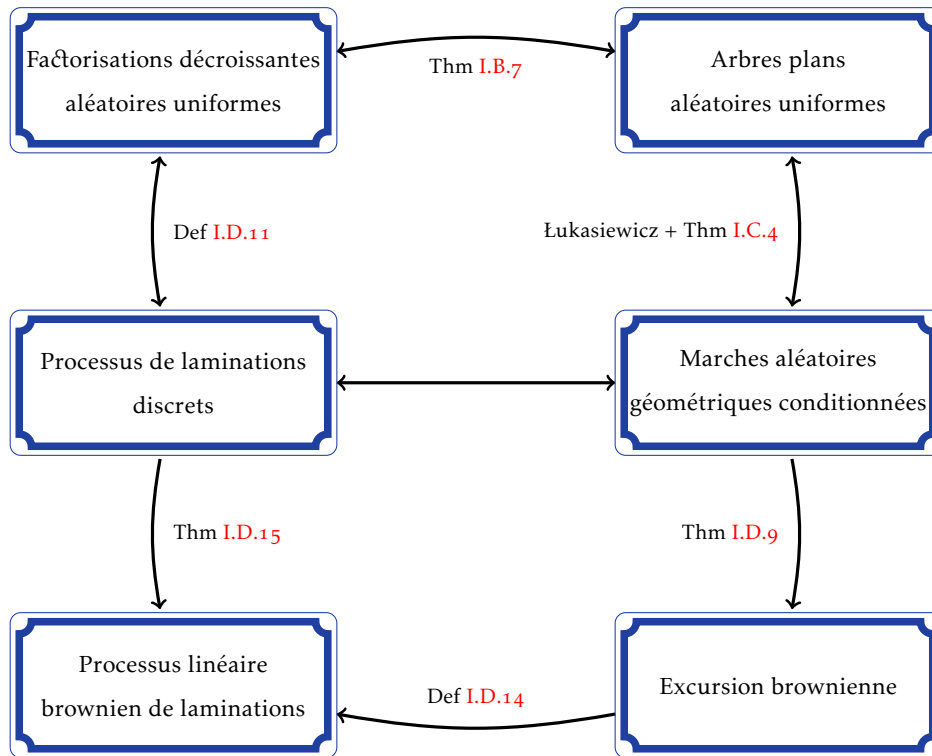


FIGURE D.10 – Approche bijective appliquée aux factorisations croissantes

Comme dit précédemment, le Théorème I.D.15 est l’analogue du Théorème 1.2 de [98] où l’auteur étudie plutôt le processus de lamination associé à une factorisation minimale uniforme. Tandis que le Théorème I.D.15 traite des factorisations décroissantes. Les deux théorèmes donnent des convergences au sens de la topologie J_1 de Skorokhod (voir, par exemple [66, Chapitre VI]). Dans le théorème 1.2 de [98], le terme de renormalisation est en \sqrt{n} ce qui signifie que les cordes macroscopiques apparaissent après un temps de l’ordre de \sqrt{n} , alors que, dans le Théorème I.D.15, le terme de renormalisation est n . Une autre différence est que, dans le cas des factorisations décroissantes, conditionnellement à ϵ , le processus de lamination $L(\epsilon)$ est déterministe tandis que dans le cas des factorisations minimales, il y a une seconde couche d’aléatoire. En effet, conditionnellement à ϵ , les cordes du processus limite sont choisies aléatoirement en jetant un nuage de points sous la courbe de ϵ d’une manière poissonnienne.

Le Théorème I.D.15 implique que, pour n’importe quelle fonctionnelle continue $F : d \rightarrow E$, où E est un espace métrique, $F(L^n)$ converge vers $F(L(\epsilon))$ en loi. Par exemple, la fonctionnelle donnant la plus grande corde actuelle est continue pour la topologie J_1 de Skorokhod (dans ce cas E est l’espace des fonctions càdlàg de $[0, 1]$

dans \mathbb{R}).

Nous pensons qu'il n'existe pas d'équivalent au Théorème I.D.15 dans le cas des factorisations croissantes. Plus précisément, si $(L_{[nt]}^n)_{0 \leq t \leq 1}$ est le processus de lamination associé à une factorisation croissante uniforme $(\tau_1^n, \dots, \tau_{n-1}^n)$, nous pensons qu'il n'y a pas de convergence en loi de ce processus au sens de la topologie J_1 de Skorokhod. En effet, en utilisant la bijection du Théorème I.B.7, on peut voir que des cordes macroscopiques vont apparaître à des temps arbitrairement proches dans L^n , pour n assez grand, ce qui rendrait la convergence J_1 impossible. Ces cordes macroscopiques apparaîtraient à chaque point de branchement de l'arbre plan associé donnant naissance, au moins, à deux sous-arbres de taille non négligeable.

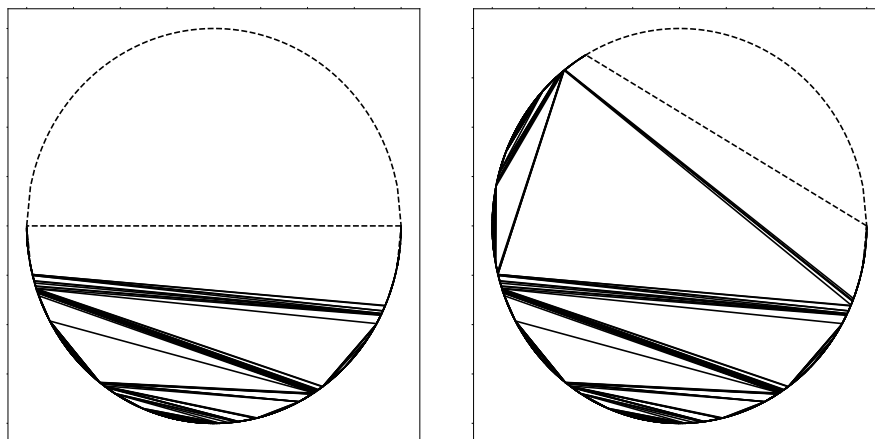


FIGURE D.11 – Représentation de $L_{1/2}(e)$ à gauche et $L_{2/3}(e)$ à droite.



CHAPITRE

II

FACTORISATIONS MINIMALES DU
CYCLE

Contenu du Chapitre

A Introduction	56
B Bijection between minimal factorizations and Cayley trees	57
B.i The bijection	57
B.ii Reformulation in terms of Cayley trees	59
C First generating function on Cayley trees	60
D A second generating function on Cayley trees and Proof of Theorem II.A.1	64
D.i A symmetry result	64
D.ii Computation of the exponential generating function of G_n	64
D.iii Proof of Theorem II.A.1	66



A Introduction

Consider the cycle permutation $(1\ 2\ \dots\ n)$ for $n \geq 2$. A *minimal factorization* is a $(n-1)$ -tuple of transpositions $(\tau_1, \dots, \tau_{n-1})$ such that $\tau_{n-1} \circ \dots \circ \tau_1 = (1\ 2\ \dots\ n)$. We denote by \mathfrak{M}_n the set of all minimal factorizations of the cycle $(1\ 2\ \dots\ n)$. Dénes [41] first showed that \mathfrak{M}_n has cardinality n^{n-2} and several bijective proofs followed afterwards (see [85], [60], [58] and [22]). Minimal factorizations are linked to other combinatorial objects such as non-crossing partitions [19] and parking functions [20] and more general factorizations have deep connections with enumerative geometry (see e.g. [7]).

Let $(\tau_1^{(n)}, \dots, \tau_{n-1}^{(n)})$ be a random minimal factorization chosen uniformly at random in \mathfrak{M}_n . The study of the behaviour of such a randomly picked minimal factorization is recent (see [51], [52] and [98]) and has a rich probabilistic structure : for instance, it is shown in [98] that such minimal factorizations have connections with Aldous-Pitman fragmentation of the Brownian continuum random tree. Here we are interested in the law of the number of times 1 and 2 appear in $(\tau_1^{(n)}, \dots, \tau_{n-1}^{(n)})$. In [52, Corollary 1.2 (iv)] it was obtained that :

$$\mathbb{P}\left(\mathbb{T}_1^{(n)} = i, \mathbb{T}_2^{(n)} = j\right) \xrightarrow{n \rightarrow \infty} e^{-2} \left[\frac{i+j-2}{(i+j-1)!} + \frac{i+j-1}{i!j!} - \frac{i+j-1}{(i+j)!} \right].$$

where $\mathbb{T}_k^{(n)} = \#\{1 \leq \ell \leq n-1 : \tau_\ell^{(n)}(k) \neq k\}$ is the number of times k appears in a transposition. We refine this result by finding explicitly the joint distribution for fixed n :

Théorème II.A.1

For $i, j \geq 1$ and $n \geq i+j$:

$$\mathbb{P}\left(\mathbb{T}_1^{(n)} = i, \mathbb{T}_2^{(n)} = j\right) = \frac{n!(n-1)^{n-i-j-1}}{(n-i-j)!(n+1)^{n-1}} \left[\frac{(i+j-2)(n-1)}{(i+j-1)!(n-i-j+1)} + \frac{i+j-1}{i!j!} - \frac{i+j-1}{(i+j)!} \right]. \quad (\text{A.1})$$

To show Theorem II.A.1, we explicitly compute the exponential generating function of the (normalized) trivariate generating function G_n defined by

$$G_n(x, y, z) = n^{n-2} \mathbb{E} \left[x^{\mathbb{T}_1^{(n)}} y^{\mathbb{T}_2^{(n)}} z^{\mathfrak{M}_1^{(n)}} \right]$$

where $\mathfrak{M}_k^{(n)} = \#\{1 \leq \ell \leq n-1 : \tau_{\ell-1}^{(n)} \circ \dots \circ \tau_1^{(n)}(k) \neq \tau_\ell^{(n)} \circ \dots \circ \tau_1^{(n)}(k)\}$ is the number of transpositions that affect the trajectory of k , and then extract the coefficient $[x^i y^j] G_n(x, y, 1)$. To this end, there are 4 main steps. First, using a known bijection between minimal factorizations and Cayley trees, we reformulate the problem in terms of a generating function of a trivariate statistic on Cayley trees (Section B). To compute this generating function, we actually start by computing another generating function F_n obtained by changing one of the three statistics (Section C). This also yields a result of independent interest by confirming a conjecture [30] involving distributional symmetries in uniform Cayley trees (Corollary II.C.2). Finally, we show bijectively that $G_n(x, y, z) = G_n(y, x, z)$ (Section D.i), and by combining this with the explicit formula of F_n we get the exponential generating function of G_n (Section D.ii) and Theorem II.A.1 follows (Section D.iii).

B Bijection between minimal factorizations and Cayley trees

B.i The bijection

Here we explain how to associate a labeled tree with a minimal factorization which will be an essential tool for us. We refer to [52] for details and proofs. Fix an integer $n \geq 2$.

Définition II.B.1

For $(\tau_1, \dots, \tau_{n-1}) \in \mathfrak{M}_n$, we define a labeled tree $\mathcal{F}(\tau_1, \dots, \tau_{n-1})$ with n vertices labeled from 1 to n where an edge labeled l is drawn between the vertices labeled a and b if and only if $\tau_l = (a, b)$ (see Figure B.1 for an example).

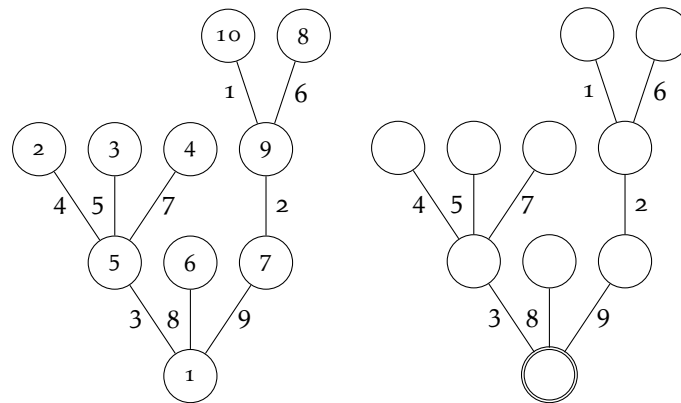


FIGURE B.1 – Representation of \mathcal{F} on the left and \mathcal{E} on the right when $n = 10$ for the minimal factorization $((9\ 10), (7\ 9), (1\ 5), (2\ 5), (3\ 5), (8\ 9), (4\ 5), (1\ 6), (1\ 7))$ of $(1 \dots 10)$. The double circle represents the root of the tree \mathcal{E} .

Clearly \mathcal{F} is injective since a minimal factorization can be easily read on its associated tree. Actually the tree \mathcal{F} gives too much information, indeed it is still possible to retrieve the associated minimal factorization when we erase the vertex-labels and keep only the edge-labels. More precisely we have :

Définition II.B.2

If $(\tau_1, \dots, \tau_{n-1}) \in \mathfrak{M}_n$, then we construct a rooted, edge-labeled tree $\mathcal{E}(\tau_1, \dots, \tau_{n-1})$ by doing the following on the tree $\mathcal{F}(\tau_1, \dots, \tau_{n-1})$:

- We root the tree at the vertex labeled 1.
- We erase all the vertex-labels (and keep only the edge-labels).

Proposition II.B.3

The map \mathcal{E} gives a bijection between the set \mathfrak{M}_n and the set \mathcal{C}'_n of rooted trees with $n - 1$ edges labeled from 1 to $n - 1$.

The set \mathcal{C}'_n is clearly in bijection with the set \mathcal{C}_n of Cayley trees with n vertices (i.e. trees with n vertices labeled from 1 to n). Indeed if $t \in \mathcal{C}_n$ we create $\alpha(t) \in \mathcal{C}'_n$ by rooting the tree t at the vertex labeled 1, then by pulling all the vertex-labels (except 1 which is erased from $\alpha(t)$) towards the root into the nearest edge. We then subtract 1 from all the labels (see Figure B.2 for an example). The map α is clearly a bijection.

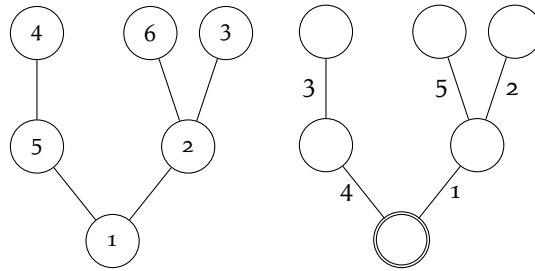


FIGURE B.2 – A tree $t \in \mathcal{C}_6$ on the left transformed into $\alpha(t) \in \mathcal{C}'_6$ on the right by pulling the vertex-labels towards the root and subtracting 1.

In particular \mathfrak{M}_n has the same cardinality as the set of Cayley trees with n vertices \mathcal{C}_n which is known to be n^{n-2} . This explains the renormalizing term in the definition of G_n . In the article [52] the authors give an explicit algorithm to find a minimal factorization back from its associated tree \mathcal{E} . We explain this algorithm right now because it will be essential for what follows.

Let $t \in \mathcal{C}'_n$ and $f \in \mathfrak{M}_n$ such that $t = \mathcal{E}(f)$. To recover f from t , we will recover $\mathcal{F}(f)$ (from which it is immediate to find f back). For this purpose, we will gradually assign the vertex-labels $1, 2, \dots, n$ to t as they can be found in $\mathcal{F}(f)$. More precisely, let us describe an algorithm `Next` that takes the tree t and an integer $k \in \{1, \dots, n - 1\}$ as arguments and assigns the vertex-label $k + 1$ to t if the vertex-label k has already been assigned in t (if k is not assigned yet `Next(t,k)` does nothing). The algorithm starts from the vertex v_0 of t labeled k . Then it follows the longest possible path of edges e_1, \dots, e_ℓ in t such that :

- e_1, \dots, e_ℓ is a *path of edges* meaning that for all $0 < i < \ell$, e_i and e_{i+1} share a common vertex v_i .
- e_1 is the edge with smallest label adjacent to v_0 .
- For all $0 < i < \ell$, e_{i+1} has the smallest label among the edges adjacent to v_i having a label greater than e_i 's label.

Starting from v_0 and following the path e_1, \dots, e_ℓ leads to the vertex of t which then gets the label $k + 1$. We also initiate the algorithm with `Next(t,0)` which gives label 1 to the root of t .

Now by applying successively `Next` to t we can label all the vertices of t in order to find back $\mathcal{F}(f)$. More precisely we have the following definition and proposition :

Définition II.B.4

For $1 \leq k \leq n$ and $t \in \mathcal{C}'_n$, we denote by $\text{Find}_k(t)$ the tree obtained by applying successively $\text{Next}(t, 0)$ then $\text{Next}(t, 1), \text{Next}(t, 2), \dots, \text{Next}(t, k-1)$ to t .

For example, applying Find_{10} to the tree on the right of Figure B.1 gives back the vertex-labels on the left of Figure B.1.

Proposition II.B.5

For $t = \mathcal{E}(f) \in \mathcal{C}'_n$:

$$\text{Find}_n(t) = \mathcal{F}(f).$$

B.ii Reformulation in terms of Cayley trees

In order to prove Theorem II.A.1, we start with rewriting the generating function G_n defined in the Introduction in terms of Cayley trees. To this purpose we introduce some notation. For A a subset of $\{1, \dots, n\}$ we denote by \mathcal{C}_A the set of trees with vertex set A . Notice that if $A = \{1, \dots, n\}$ then $\mathcal{C}_A = \mathcal{C}_n$ is the set of *Cayley trees* with n vertices. If $t \in \mathcal{C}_A$ and $i \in A$, we denote by $\text{deg}_i(t)$ the degree of the vertex i in the tree t . Also, similarly to the algorithm Next (introduced in the previous section) we consider a certain path in t starting from i , namely :

Définition II.B.6

For $t \in \mathcal{C}_A$ and $i \in A$ we consider the longest path of vertices, starting from i , such that each vertex of the path has the smallest label among the ones that are both adjacent to the previous vertex on the path and have a greater label than the label of the previous vertex on the path. We denote by $L_i(t)$ the set of vertices composing this path but without including the first one which is the vertex i (see Figure B.3 for an example).

Définition II.B.7

For $t \in \mathcal{C}_n$ we denote by $\text{deg}'_2(t)$ the degree, in t , of the last vertex of the path $L_1(t)$ (see Figure B.3 for an example).

Recall from Section B.i the bijection \mathcal{E} between \mathfrak{M}_n and \mathcal{C}'_n as well as a bijection α from \mathcal{C}_n to \mathcal{C}'_n (illustrated in Figure B.2). Notice that if $f \in \mathfrak{M}_n$ and $t = \alpha \circ \mathcal{E}(f)$ then

$$(\mathbb{T}_1(f), \mathbb{T}_2(f), \mathbb{M}_1(f)) = (\text{deg}_1(t), \text{deg}'_2(t), |L_1(t)|).$$

where $\mathbb{T}_i(f)$ is the number of times i appears in f and $\mathbb{M}_j(f)$ is the number of transpositions in f that affect the



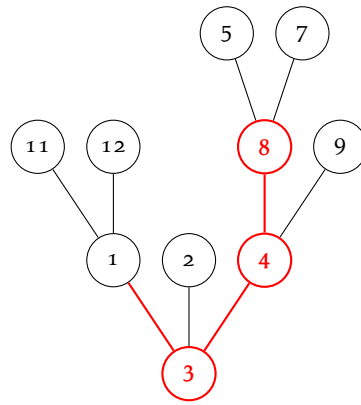


FIGURE B.3 – Example of a tree $t \in \mathcal{C}_A$ with $A = \{1, 2, 3, 4, 5, 7, 8, 9, 11, 12\}$. The path $L_1(t) = \{3, 4, 8\}$ starting from 1 is represented in thick red. Here $\text{deg}'_2(t) = \text{deg}_8(t) = 3$.

trajectory of j . The last identity allows us to reformulate the definition of G_n for $n \geq 2$:

$$G_n(x, y, z) = \sum_{t \in \mathcal{C}_n} x^{\text{deg}_1(t)} y^{\text{deg}'_2(t)} z^{|L_1(t)|}. \tag{B.1}$$

In the next section, in order to compute G_n , we introduce another generating function F_n whose definition is similar to (B.1) except that deg'_2 is replaced with deg_2 .

C First generating function on Cayley trees

For $n \geq 2$ we define the generating function :

$$F_n(x, y, z) = \sum_{t \in \mathcal{C}_n} x^{\text{deg}_1(t)} y^{\text{deg}_2(t)} z^{|L_1(t)|}.$$

By convention we also set for $n \geq 1$:

$$F_n(x, 1, 1) = F_n(x) = \sum_{t \in \mathcal{C}_n} x^{\text{deg}_1(t)} = x(n-1+x)^{n-2}.$$

The last equality is a well known result on Cayley trees. Actually, it turns out we have an explicit formula for $F_n(x, y, z)$.

Proposition II.C.1

For $n \geq 3$,

$$F_n(x, y, z) = xyz \left[x(n-2+x)^{n-3} \left(1 - \frac{yz}{z+y-1} \right) + (x+y+z+n-3)^{n-3} \left(n-2+yz + \frac{xyz}{z+y-1} \right) \right]. \tag{C.1}$$

Formula (C.1) still holds in the cases $(y = 1, n = 2)$ and $(y = z = 1, n = 1)$. Proposition II.C.1 implies that F_n is symmetric in y and z thus we have the following corollary which confirms a conjecture made by Caraceni [30] :

Corollaire II.C.2

Let T_n be a random uniform Cayley tree with n vertices, then the two triplets $(\deg_1(T_n), |L_1(T_n)|, \deg_2(T_n))$ and $(\deg_1(T_n), \deg_2(T_n), |L_1(T_n)|)$ have the same law.

It would be very interesting to obtain a direct bijective proof of Corollary II.C.2. Formula (C.1) with $y = 1$ was first conjectured in [52, Conjecture 1.4] and was proved by O. Angel & J. Martin [11]. Below, we give Angel and Martin’s proof of the case $y = 1$ which will be useful to deduce the general case.

Preuve de la Proposition II.C.1 for $y = 1$

We show by induction on $n \geq 2$ that $F_n(x, 1, z) = xzf_n(x + z)$ where f_n is a real-valued function. It is obviously true for $n = 2$ with $f_2 = 1$. Suppose that is true for all $2 \leq k \leq n - 1$ with $n \geq 3$.

We denote by \mathcal{P}_n the set of couples (A, B) with A and B two subsets of $\{1, \dots, n\}$ such that $A \cup B = \{1, \dots, n\}$, $A \cap B = \emptyset$, $1 \in A$ and $n \in B$. Let $t \in \mathcal{C}_n$. Set the vertex 1 to be the root of t . Consider cutting the tree t by removing the edge between the vertex n and its parent to end up with two trees $t_1 \in \mathcal{C}_A$ and $t_2 \in \mathcal{C}_B$ for some $(A, B) \in \mathcal{P}_n$. Actually given two trees $t_1 \in \mathcal{C}_A$ and $t_2 \in \mathcal{C}_B$ with $(A, B) \in \mathcal{P}_n$ there are $|A|$ distinct ways to attach t_2 to t_1 by joining the vertex $n + 1$ of t_2 to one of t_1 ’s vertices to obtain a tree $t \in \mathcal{C}_n$. With this in mind we can decompose the quantity $F_n(x, 1, z)$ depending on where t_2 is attached to t_1 :

$$F_n(x, 1, z) = xz + \sum_{\substack{(A,B) \in \mathcal{P}_n \\ |A| > 1}} \sum_{\substack{t_1 \in \mathcal{C}_A \\ t_2 \in \mathcal{C}_B}} (|A| - 2)x^{\deg_1(t_1)}z^{|L_1(t_1)|} + x^{\deg_1(t_1)+1}z^{|L_1(t_1)|} + x^{\deg_1(t_1)}z^{|L_1(t_1)|+1}. \quad (C.2)$$

The first term corresponds to the case where t_1 has only 1 vertex (i.e. $|A| = 1$). If $|A| > 1$ then the vertex 1 and the last vertex of the path $L_1(t_1)$ are distinct in t_1 and we have to consider three cases :


- The tree t_2 is attached to a vertex of t_1 which is neither 1, nor the last vertex of $L_1(t_1)$.
- The tree t_2 is attached to the vertex 1.
- The tree t_2 is attached to the last vertex of $L_1(t_1)$.

We then have :

$$F_n(x, 1, z) = xz + \sum_{a=2}^{n-1} \binom{n-2}{a-1} (n-a)^{n-a-2} F_a(x, 1, z)(a-2+x+z).$$

By induction we conclude that :

$$F_n(x, 1, z) = xz + xz \sum_{a=2}^{n-1} \binom{n-2}{a-1} (n-a)^{n-a-2} f_a(x+z)(a-2+x+z).$$

So $F_n(x, 1, z)/(xz)$ depends only on $x + z$, thus induction is shown. We then just need to take $z = 1$ and use the case $y = z = 1$ to conclude. 

To prove the general case of Proposition II.C.1 we will use the particular cases $y = 1$ and $z = 1$. We will also use the following Abel’s binomial theorem [90, p. 18].

Proposition II.C.3

For every integer $n \geq 0$ the following identity holds :

$$\sum_{k=0}^n \binom{n}{k} x(x-kz)^{k-1} (y+kz)^{n-k} = (x+y)^n.$$

Three useful variants can be deduced from this identity.

Corollaire II.C.4

For every integer $n \geq 0$,

Variant 1 $\sum_{k=0}^n \binom{n}{k} (x+k)^{k-1} (y-k)^{n-k} = \frac{(x+y)^n}{x};$

Variant 2 $\sum_{k=0}^n \binom{n}{k} (x+k)^k (n-k+y)^{n-k-1} = \frac{(x+y+n)^n}{y};$

Variant 3 $\sum_{k=0}^n \binom{n}{k} (x+k)^{k-1} (n-k+y)^{n-k-1} = \frac{x+y}{xy} (x+y+n)^{n-1}.$

Preuve

Taking $z = -1$ in Abel's binomial formula gives the first variant. Doing the change of index $k \rightarrow n - k$ and the change of variables $y \rightarrow x + n$ and $x \rightarrow y$ in variant 1 gives variant 2. To get variant 3 we begin by differentiating variant 1 with respect to y , so we have :


$$\sum_{k=0}^n \binom{n}{k} (n-k)(x+k)^{k-1} (n-k+y)^{n-k-1} = n \frac{(x+y+n)^{n-1}}{x}.$$

Denote by A the left side of variant 3 which we want to compute, then :

$$nA - n \sum_{k=1}^n \binom{n-1}{k-1} (x+k)^{k-1} (n-k+y)^{n-k-1} = n \frac{(x+y+n)^{n-1}}{x}.$$

By doing the change of index $k \rightarrow k + 1$ in the last sum and using variant 2 for the resulting sum we get :

$$nA - n \frac{(x+y+n)^{n-1}}{y} = n \frac{(x+y+n)^{n-1}}{x}.$$

The expression of A can be deduced from the last display and thus variant 3 is shown. 

Preuve de la Proposition II.C.1 in the general case

Assume $n \geq 3$. Once again we will use a "tree-cutting" argument but instead of cutting at vertex n , we cut at vertex 2. More precisely, we denote by \mathcal{Q}_n the set of all couples (A, B) with A and B two subsets of $\{1, \dots, n\}$ such that $A \cup B = \{1, \dots, n\}$, $A \cap B = \emptyset$, $1 \in A$ and $2 \in B$. Once again, we decompose the quantity $F_n(x, y, z)$ depending on where t_2 is attached to t_1 :

$$F_n(x, y, z) = \sum_{(A,B) \in \mathcal{Q}_n} \sum_{\substack{t_1 \in \mathcal{C}_A \\ t_2 \in \mathcal{C}_B}} \left[(|A|-1)x^{\deg_1(t_1)}y^{\deg_2(t_2)+1}z^{|L_1(t_1)|} + x^{\deg_1(t_1)+1}y^{\deg_2(t_2)+1}z^{|L_2(t_2)|+1} \right].$$

The first term appearing after the sums corresponds to attaching t_2 to a vertex which is not 1 in t_1 and the second term corresponds to attaching t_2 to the vertex 1. We then have :


$$F_n(x, y, z) = \sum_{a=2}^{n-1} \binom{n-2}{a-1} (a-1)yF_a(x, 1, z)F_{n-a}(y) + \sum_{a=1}^{n-1} \binom{n-2}{a-1} xyzF_a(x)F_{n-a}(y, 1, z).$$

Now we can use the cases $y = 1$ and $y = z = 1$ to replace the occurrences of F_n in the last display. Let's compute the first sum which we call $A_n(x, y, z)$, afterwards we will compute the second one, $B_n(x, y, z)$. We have :

$$\begin{aligned} A_n(x, y, z) &= \sum_{a=2}^{n-1} \binom{n-2}{a-1} (a-1)xy^2z(a-2+x+z)^{a-2}(n-a-1+y)^{n-a-2} \\ &= (n-2) \sum_{a=0}^{n-3} \binom{n-3}{a} xy^2z(a+x+z)^a(n-a-3+y)^{n-a-4} \\ &= (n-2)xyz(x+y+z+n-3)^{n-3}. \end{aligned}$$

The second equality comes from the fact that $(a-1)\binom{n-2}{a-1} = (n-2)\binom{n-3}{a-2}$. The last equality comes from variant 2 of Abel's binomial formula. Now for $B_n(x, y, z)$ we need to be careful and isolate the case $a = n-1$ because Formula (C.1) doesn't apply in the case $(y = 1, n = 1)$. We get :

$$\begin{aligned} B_n(x, y, z) - x^2yz(x+n-2)^{n-3} &= \sum_{a=0}^{n-3} \binom{n-2}{a} x^2y^2z^2(a+x)^{a-1}(n-a-3+y+z)^{n-a-3} \\ &= \sum_{a=0}^{n-4} \binom{n-3}{a} x^2y^2z^2(a+1+x)^a(n-a-4+y+z)^{n-a-4} \\ &\quad + \sum_{a=0}^{n-3} \binom{n-3}{a} x^2y^2z^2(a+x)^{a-1}(n-a-3+y+z)^{n-a-3}. \\ &= \frac{x^2y^2z^2}{y+z-1} \left[(x+y+z+n-3)^{n-3} - (x+n-2)^{n-3} \right] \\ &\quad + xy^2z^2(x+y+z+n-3)^{n-3}. \end{aligned}$$

The second equality comes from the fact that $\binom{n-2}{a} = \binom{n-3}{a-1} + \binom{n-3}{a}$. The last one comes from variants 1 and 2 of Abel's binomial theorem. The desired formula follows easily. 

D A second generating function on Cayley trees and Proof of Theorem

II.A.1

The goal now is to compute the exponential generating function of G_n :

$$\sum_{n \geq 1} \frac{G_{n+1}(x, y, z)}{n!} t^n.$$

Then by identifying coefficients in Formula (D.2), we will be able to prove Theorem II.A.1. The first step is to establish a symmetry property of G_n with a bijective approach.

D.i A symmetry result


Before computing the exponential generating function of G_n we first state a useful symmetry result.

Proposition II.D.1

For all $n \geq 2$:

$$G_n(x, y, z) = G_n(y, x, z).$$

Preuve

We will prove it by finding a bijection ϕ in \mathfrak{M}_n which exchanges \mathbb{T}_1 and \mathbb{T}_2 and keeps \mathbb{M}_1 unchanged. For $1 \leq k \leq n$ we set $\gamma(k) = 3 - k \pmod n$ so γ is a permutation and $\gamma^{-1} = \gamma$. For $(\tau_1, \dots, \tau_{n-1}) \in \mathfrak{M}_n$ we define $\phi(\tau_1, \dots, \tau_{n-1}) = (\gamma \circ \tau_{n-1} \circ \gamma, \dots, \gamma \circ \tau_1 \circ \gamma)$. Notice that for any transposition $\tau = (a b)$, $\gamma \circ \tau \circ \gamma = (\gamma(a) \gamma(b))$ hence $\phi(\tau_1, \dots, \tau_{n-1})$ is a $(n - 1)$ -tuple of transpositions. To see that ϕ has the expected property, we interpret the action of ϕ on the tree \mathcal{F} . The tree $\mathcal{F} \circ \phi$ is obtained from \mathcal{F} by relabelling the vertex-labels according to the permutation γ and the edge-labels according to the permutation $i \mapsto n - i$. Such an edge-relabelling implies that Find_n reads through $\mathcal{E} \circ \phi$ in the exact opposite order than in \mathcal{E} . Thus we easily check that $\phi(\tau_1, \dots, \tau_{n-1})$ sends 2 on 3, 3 on 4, ..., n on 1 and 1 on 2 so $\phi(\tau_1, \dots, \tau_{n-1})$ is a minimal factorization of $(1 \dots n)$. 

D.ii Computation of the exponential generating function of G_n

Before computing the exponential generating function of G_n let's introduce the Lambert W function (see e.g. [38]). It is by definition the solution (in the sense of formal series) of $W(z)e^{W(z)} = z$. Using Lagrange inversion, one can show that $W(z) = \sum_{n \geq 1} \frac{(-n)^{n-1}}{n!} z^n$. Again, using Lagrange inversion, one can also compute :

$$e^{-rW(-z)} = \left[\frac{-W(-z)}{z} \right]^r = \sum_{n \geq 0} \frac{r(n+r)^{n-1}}{n!} z^n. \tag{D.1}$$

These properties of W will be useful when proving Proposition II.D.2 and Theorem II.A.1.

Proposition II.D.2

The following identity on formal series holds :

$$e^t(y-x) \sum_{n \geq 1} \frac{G_{n+1}(x, y, z)}{n!} t^n = \frac{xyz(x-1)}{x+z-1} e^{-yW(-t)} - \frac{xyz(y-1)}{y+z-1} e^{-xW(-t)} + \frac{xyz^2(y-x)}{(x+z-1)(y+z-1)} e^{-(x+y+z-1)W(-t)}. \quad (\text{D.2})$$

Preuve

Fix $n \geq 1$. As in the proof of Proposition II.C.1 in the case $y = 1$, we denote by \mathcal{P}_{n+1} the set of couples (A, B) with A and B two subsets of $\{1, \dots, n+1\}$ such that $A \cup B = \{1, \dots, n+1\}$, $A \cap B = \emptyset$, $1 \in A$ and $n+1 \in B$. Then with a similar "tree-cutting" argument that led to (C.2) we have :

$$G_{n+1}(x, y, z) = xyz \sum_{t \in \mathcal{C}_n} y^{\deg_1(t)} + \sum_{\substack{(A,B) \in \mathcal{P}_{n+1} \\ |A| > 1}} \sum_{\substack{t_1 \in \mathcal{C}_A \\ t_2 \in \mathcal{C}_B}} \left[(|A|-2)x^{\deg_1(t_1)}y^{\deg'_2(t_1)}z^{|L_1(t_1)|} + x^{\deg_1(t_1)+1}y^{\deg'_2(t_1)}z^{|L_1(t_1)|} + x^{\deg_1(t_1)}y^{\deg_{n+1}(t_2)+1}z^{|L_1(t_1)|+1} \right].$$

The first term corresponds to the case where t_1 has only 1 vertex (i.e. $|A| = 1$). If $|A| > 1$ then the vertex 1 and the last vertex of the path $L_1(t_1)$ are distinct in t_1 and we have to consider three cases :

- We attach t_2 to a vertex of t_1 which is neither 1, nor the last vertex of $L_1(t_1)$.
- We attach t_2 to the vertex 1.
- We attach t_2 to the last vertex of $L_1(t_1)$.

We then have :

$$G_{n+1}(x, y, z) = xyzF_n(y) + \sum_{a=2}^n \binom{n-1}{a-1} ((a-2)G_a(x, y, z) + xG_a(x, y, z) + yzF_a(x, 1, z)F_{n+1-a}(y)).$$

By Proposition II.D.1, we get :

$$G_{n+1}(x, y, z) = xyzF_n(x) + \sum_{a=2}^n \binom{n-1}{a-1} [(a-2)G_a(x, y, z) + yG_a(x, y, z) + xzF_a(y, 1, z)F_{n+1-a}(x)].$$

Now if we make the difference between the last two equations, we obtain :

$$0 = xyz(F_n(y) - F_n(x)) + (x-y) \sum_{a=2}^n \binom{n-1}{a-1} G_a(x, y, z) + z \sum_{a=2}^n \binom{n-1}{a-1} [yF_a(x, 1, z)F_{n+1-a}(y) - xF_a(y, 1, z)F_{n+1-a}(x)].$$

If $n \geq 2$, by using the third variant of Abel's binomial formula and Proposition II.C.1 we get :

$$\sum_{a=2}^n \binom{n-1}{a-1} F_a(x, 1, z)F_{n+1-a}(y) = xz \frac{x+y+z-1}{x+z-1} (x+y+z+n-2)^{n-2} - \frac{xyz(x+n-1)^{n-2}}{y+z-1}$$



and

$$\sum_{a=2}^n \binom{n-1}{a-1} F_a(y, 1, z) F_{n+1-a}(x) = yz \frac{x+y+z-1}{y+z-1} (x+y+z+n-2)^{n-2} - \frac{xyz(y+n-1)^{n-2}}{x+z-1}.$$

So, finally,

$$0 = (x-y) \sum_{a=2}^n \binom{n-1}{a-1} G_a(x, y, z) + u_{n-1}(x, y, z),$$

where

$$\begin{aligned} u_{n-1}(x, y, z) &= \frac{xy^2z(x-1)}{x+z-1} (y+n-1)^{n-2} - \frac{x^2yz(y-1)}{y+z-1} (x+n-1)^{n-2} \\ &+ \frac{xyz^2(y-x)(x+y+z-1)}{(x+z-1)(y+z-1)} (x+y+z+n-2)^{n-2}. \end{aligned}$$

Thus, by Pascal's inversion formula,


$$(y-x)G_{n+1}(x, y, z) = \sum_{k=0}^n \binom{n}{k} (-1)^{n-k} u_k(x, y, z). \tag{D.3}$$

Now define for $n \geq 0$ the polynomials P_n by

$$P_n(u) = \sum_{k=0}^n u \binom{n}{k} (-1)^{n-k} (k+u)^{k-1}. \tag{D.4}$$

Their exponential generating function is given by

$$\begin{aligned} \sum_{n \geq 0} \frac{P_n(u)}{n!} t^n &= \sum_{k \geq 0} u t^k (k+u)^{k-1} \sum_{n \geq k} \binom{n}{k} \frac{(-1)^{n-k}}{n!} t^{n-k} \\ &= u e^{-t} \sum_{k \geq 0} \frac{(k+u)^{k-1}}{k!} t^k \\ &= e^{-t-uW(-t)}. \end{aligned}$$

This, combined with (D.3), readily gives the desired result. 

D.iii Proof of Theorem II.A.1

To simplify notation, for $n \geq 2$, set

$$p_{i,j}^n = \mathbb{P}(\deg_1(T_n) = i, \deg_2'(T_n) = j)$$

so that

$$p_{i,j}^n = \frac{1}{n^{n-2}} [x^i y^j] G_n(x, y, 1).$$

Preuve du Théorème II.A.1

Fix $i, j \geq 1$. We take the formula of Proposition II.D.2 with $z = 1$ and divide it by $(y-x)$ to get :

$$\begin{aligned} e^t \sum_{n \geq 1} \frac{G_{n+1}(x, y, 1)}{n!} t^n &= \frac{xy}{y-x} [e^{-yW(-t)} - e^{-xW(-t)}] + \frac{1}{y-x} [xe^{-xW(-t)} - ye^{-yW(-t)}] \\ &+ e^{-(x+y)W(-t)}. \end{aligned}$$

We shall identify the coefficient in front of $x^i y^j$ (which is a polynomial in t) in this formula. On the left of this equality, the coefficient is :

$$e^t \sum_{n \geq 1} p_{i,j}^{n+1} \frac{(n+1)^{n-1}}{n!} t^n.$$

Using the identity $y^k - x^k = (y-x)(y^{k-1} + y^{k-2}x + \dots + x^{k-1})$ we deduce the coefficient on the right :

$$\frac{(-W(-t))^{i+j}}{i!j!} + \frac{(-W(-t))^{i+j-1}}{(i+j-1)!} + \frac{(-W(-t))^{i+j}}{(i+j)!}.$$

Therefore

$$\sum_{n \geq 1} p_{i,j}^{n+1} \frac{(n+1)^{n-1}}{n!} t^n = e^{-t} \frac{(-W(-t))^{i+j}}{i!j!} + e^{-t} \frac{(-W(-t))^{i+j-1}}{(i+j-1)!} + e^{-t} \frac{(-W(-t))^{i+j}}{(i+j)!}.$$

We now fix $n \geq i+j$ and identify the coefficient associated with t^n in the above formula. Fix $1 \leq \ell \leq n$.

By using Formula (D.1) we obtain :

$$[t^n] e^{-t} (-W(-t))^\ell = \ell \sum_{k=\ell}^n \frac{(-1)^{n-k}}{(n-k)!} \frac{k^{k-1-\ell}}{(k-\ell)!}.$$

Recall the definition of P_n in (D.4). Taking $y = z = 1$ in Equation (D.3) gives :

$$G_{n+1}(x, 1, 1) = F_{n+1}(x) = P_n(x+1) - P_n(1).$$

In particular,


$$P_n^{(\ell)}(x+1) = F_n^{(\ell)}(x) = x \frac{(n-1)!}{(n-1-\ell)!} (n+x)^{n-1-\ell} + \ell \frac{(n-1)!}{(n-\ell)!} (n+x)^{n-\ell},$$

with the convention that $1/(-1)! = 1/\Gamma(0) = 0$. On the other hand, by definition of P_n ,

$$P_n^{(\ell)}(x) = x \sum_{k=\ell+1}^n (-1)^{n-k} \binom{n}{k} \frac{(k-1)!}{(k-1-\ell)!} (k+x)^{k-1-\ell} + \ell \sum_{k=\ell}^n (-1)^{n-k} \binom{n}{k} \frac{(k-1)!}{(k-\ell)!} (k+x)^{k-\ell}.$$

We finally obtain :

$$\ell \sum_{k=\ell}^n \frac{(-1)^{n-k}}{(n-k)!} \frac{k^{k-1-\ell}}{(k-\ell)!} = \frac{1}{n!} P_n^{(\ell)}(0) = (n-1)^{n-\ell-1} \frac{\ell-1}{(n-\ell)!}.$$

Formula (A.1) then follows. 



CHAPITRE

III

TUPLES DE STATIONNEMENT

Contenu du Chapitre

A Introduction	70
B Bijection between parking functions and Cayley trees	72
C Convergence for the total variation distance	74
C.i Probability that the parking places are distinct	74
C.ii A monotonicity argument	75
C.iii Bounding the moments of a random walk conditioned to be an excursion	77
C.iv Proof of Theorem III.A.1 (i)	81
D Convergence for the Kolmogorov distance	81
E Sum and maximum of the first parking places	84



A Introduction

A *parking function* of size n is a function $\pi_n : \llbracket 1, n \rrbracket \rightarrow \llbracket 1, n \rrbracket$ such that, if $\pi'_n(1) \leq \dots \leq \pi'_n(n)$ is the nondecreasing rearrangement of $(\pi_n(1), \dots, \pi_n(n))$, then $\pi'_n(i) \leq i$ for all i . Konheim and Weiss [75] first introduced parking functions, in a context of information storing, to study hashing functions and they have shown that there are $(n + 1)^{n-1}$ parking functions of size n . Since then, parking functions became a subject of interest in the fields of combinatorics, probability, group theory and computer science. More precisely, parking functions are linked to the enumerative theory of trees and forests [35], to coalescent processes [34, 28], to the analysis of set partitions [94], hyperplane arrangements [95, 92], polytopes [96, 36] and sandpile groups [37]. Finally, the study of probabilistic properties of parking functions has recently attracted some interest [45, 73, 104]. We refer to [103] for an extensive survey. Our initial interest for parking functions comes from the study of minimal factorisations of cycles [20].

For all $n \geq 1$ consider a random parking function $(\pi_n(i))_{1 \leq i \leq n}$ chosen uniformly among all the $(n + 1)^{n-1}$ possible parking functions of size n . For all $1 \leq k \leq n$ denote by

$$d_{TV}(k, n) := \frac{1}{2} \sum_{i_1, \dots, i_k=1}^n \left| \mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) - \frac{1}{n^k} \right| \tag{A.1}$$

the total variation distance between $(\pi_n(1), \dots, \pi_n(k))$ and $(U_n(1), \dots, U_n(k))$ where $(U_n(i))_{1 \leq i \leq n}$ are i.i.d. uniformly distributed in $\llbracket 1, n \rrbracket$. Diaconis and Hicks [45, Corollary 6] have shown that $d_{TV}(1, n)$ tends to 0 as n tends to infinity and conjectured that for any fixed k , $d_{TV}(k, n)$ should be a $O(k/\sqrt{n})$. In the same paper the authors studied the Kolmogorov distance

$$d_K(k, n) := \max_{1 \leq i_1, \dots, i_k \leq n} \left| \mathbb{P}(\pi_n(1) \leq i_1, \dots, \pi_n(k) \leq i_k) - \frac{i_1 \dots i_k}{n^k} \right| \tag{A.2}$$

and have shown that [45, Theorem 3] for $1 \leq k \leq n$:

$$d_K(k, n) = O\left(k\sqrt{\frac{\log n}{n}} + \frac{k^2}{n}\right).$$

They also discuss the growth threshold of k at which d_K doesn't converge towards 0 anymore and find that for k of order n the convergence fails. We prove a stronger version of Diaconis' and Hicks' conjecture when k is allowed to grow with n at rate at most \sqrt{n} . Moreover the Kolmogorov distance converges towards 0 when $k = o(n)$, namely :

Théorème III.A.1

(i) If $k_n = o(\sqrt{n})$ then

$$d_{TV}(k_n, n) = O\left(\frac{k_n}{\sqrt{n}}\right). \tag{A.3}$$

(ii) If $k_n = o(n)$ and $\sqrt{n} = o(k_n)$ then

$$d_K(k_n, n) = O\left(\frac{\sqrt{n}}{k_n} + \left(\frac{k_n}{n}\right)^{0.19}\right). \tag{A.4}$$

Remarque

In Theorem III.A.1 (ii), \sqrt{n} is assumed to be a $o(k_n)$. Since the function $k \mapsto d_K(k, n)$ is nondecreasing for fixed n , the distance $d_K(k_n, n)$ still tends towards 0 as long as $k_n = o(n)$. Thus, sequence $a_n = n$ satisfies $d_K(k_n, a_n) \rightarrow 0$ if $k_n = o(a_n)$ and $d_K(k_n, a_n) \not\rightarrow 0$ if $a_n = O(k_n)$. It would be very interesting to identify such a sequence for d_{TV} instead of d_K , and, in particular, to see if $d_{TV}(k_n, n) \rightarrow 0$ when $k_n = o(n)$.

The main idea to prove Theorem III.A.1 is to express the law of π_n in terms of a conditioned random walk (Proposition III.B.3 below) and to control its moments, uniformly in time (Proposition III.C.4). Such uniform estimates on conditioned random walks are delicate to establish, and we believe them to be of independent interest. As an application, we obtain limit theorems for the maximum and the sum of the first k_n parking places. Namely we obtain the following corollary (whose proof is postponed to the last section) :

Corollaire III.A.2

(i) If $k_n = o(\sqrt{n})$ and $k_n \rightarrow \infty$ then the convergence

$$\sqrt{\frac{12}{k_n}} \left(\frac{\pi_n(1) + \dots + \pi_n(k_n)}{n} - \frac{k_n}{2} \right) \rightarrow \mathcal{N}(0, 1)$$

holds in distribution where $\mathcal{N}(0, 1)$ is a standard normal distribution.

(ii) If $k_n = o(n)$ and $k_n \rightarrow \infty$ then the convergence

$$k_n \left(1 - \frac{1}{n} \max\{\pi_n(1), \dots, \pi_n(k_n)\} \right) \rightarrow \mathcal{E}(1)$$

holds in distribution where $\mathcal{E}(1)$ is an exponential distribution with mean 1.

Remarque

The complete sum $\pi_n(1) + \dots + \pi_n(n)$ has been studied and converges, after renormalization, towards a more complicated distribution involving zeros of the Airy function (see [45, Theorem 14]).

When $k_n \sim cn$ we obtain the following limit theorem for the first k_n parking places. The proof uses other techniques and Proposition III.B.3 (or rather its proof).

Proposition III.A.3

If $k_n \sim cn$ with $c \in (0, 1]$ then for all $a \in \mathbb{N}$ there exists a integer-valued random variable S_a^* such that $0 \leq S_a^* \leq a$ almost surely and

$$\mathbb{P}(n - \max\{\pi_n(1), \dots, \pi_n(k_n)\} \geq a) \rightarrow \mathbb{E}[(1 - c)^{a - S_a^*}].$$

In section B we use a bijection between parking functions and Cayley trees and use it to reformulate the law of π_n in terms of conditioned random walks. Then in section C we bound the moments of a conditioned random walk in order to control the probability mass function of π_n and prove Theorem III.A.1 (i). In section D we prove (ii) using arguments developed in the previous sections. Finally the last section is devoted to the



proof of Corollary III.A.2 and Proposition III.A.3.

From now until the end of this chapter, C denotes a constant which may vary from line to line.

B Bijection between parking functions and Cayley trees

Here the goal is to use the bijection found by Chassaing and Marckert in [35] between parking functions of size n and Cayley trees with $n + 1$ vertices (i.e. acyclic connected graphs with $n + 1$ vertices labeled from 0 to n). This bijection will allow to express the joint distribution of the first positions of a uniform parking function in terms of random walks. To this end, we start with some notation and definitions. Let \mathcal{C}_{n+1} be the set of Cayley trees with $n + 1$ vertices labeled from 0 to n where the vertex labeled 0 is distinguished from the others (we call it the *root* of the tree). Also let P_n be the set of parking functions of size n . We consider the *breadth first search* on a tree $t \in \mathcal{C}_{n+1}$ by ordering the children of each vertex of t in the increasing order of their labels (thus t is viewed as a plane tree) and then taking the regular breadth first search associated to the plane order (see [35] for a detailed definition and see figure B.1 for an example). For $t \in \mathcal{C}_{n+1}$ and $1 \leq i \leq n$, define $r(i, t)$ to be the rank for the breadth first search of the parent of the vertex labeled i in t . The bijection of Chassaing and Marckert is described in the following theorem.

Théorème III.B.1 : Chassaing and Marckert [35]

The map

$$t \mapsto (r(1, t), \dots, r(n, t)) \tag{B.1}$$

is a bijection between \mathcal{C}_{n+1} and P_n .

Remarque

Chassaing and Louchard [34] described a similar bijection using what they call the *Standard* order instead of the breadth first search.

Let $(X_i)_{i \geq 1}$ be independent and identically distributed random variables distributed as a Poisson distribution of parameter 1. For all $n \geq 0$ we set $S_n := \sum_{i=1}^n (X_i - 1)$ and for all $a \in \mathbb{Z}$, $\tau_a := \min\{n \geq 1 : S_n = a\}$ the first time that the random walk $(S_n)_n$ reaches a . Consider the probability measure $\mathbb{P}_n := \mathbb{P}(\cdot \mid \tau_{-1} = n + 1)$ and set $\mathbb{E}_n := \mathbb{E}[\cdot \mid \tau_{-1} = n + 1]$. It is well known that a Bienaymé-Galton-Watson tree with a critical Poisson reproduction law conditioned on having n vertices has the same distribution, when we uniformly randomly label the vertices from 1 to n , as a uniform Cayley tree with n vertices (see e.g. [68, Example 10.2]). From this, Chassaing and Marckert deduce the following Corollary.

Corollaire III.B.2 : Chassaing and Marckert [35]

Let T_{n+1} be a random Cayley tree in \mathcal{C}_{n+1} with uniform distribution. Under \mathbb{P}_n , the random vector $(\#\{1 \leq j \leq n : r(j, T_{n+1}) = i\})_{1 \leq i \leq n+1}$ has the same distribution as $(X_i)_{1 \leq i \leq n+1}$.

We are now able to state and prove the main result of this section.

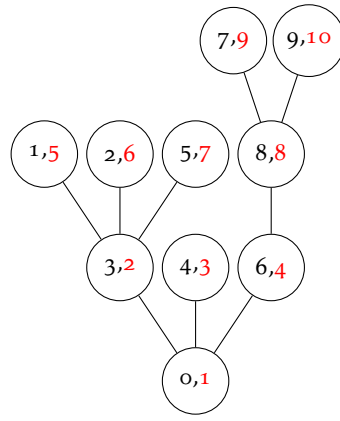


FIGURE B.1 – Example of a Cayley tree t with 10 vertices. For every vertex, its label is represented in black on the left and its rank for the breadth first search is in red on the right. For instance, here we have $r(5, t) = 2$. The parking function associated with this tree by (B.1) is $(2, 2, 1, 1, 2, 1, 8, 4, 8)$.

Proposition III.B.3

Fix $1 \leq k \leq n$ and $1 \leq i_1, \dots, i_k \leq n$. Let $j_1 < \dots < j_r$ be such that $\{i_1, \dots, i_k\} = \{j_1, \dots, j_r\}$ (so the vector (j_1, \dots, j_r) is the vector (i_1, \dots, i_k) where repeated elements have been reduced to a single one, and r is the number of distinct elements of (i_1, \dots, i_k)). Define $m_s = \#\{u : i_u = j_s\}$ for all $1 \leq s \leq r$. Then


$$\mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) = \frac{(n-k)!}{n!} \mathbb{E}_n \left[\prod_{s=1}^r (X_{j_s})_{m_s} \right] \tag{B.2}$$

where $(x)_m := x(x-1)\dots(x-m+1)$.

Preuve

Let T_{n+1} be a random Cayley tree in \mathcal{C}_{n+1} with uniform distribution. Denote by $\mathfrak{S}(k, n)$ the set of all injections between $\llbracket 1, k \rrbracket$ and $\llbracket 1, n \rrbracket$. We have

$$\begin{aligned} \mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) &= \frac{(n-k)!}{n!} \sum_{\sigma \in \mathfrak{S}(k, n)} \mathbb{P}(\pi_n(\sigma(1)) = i_1, \dots, \pi_n(\sigma(k)) = i_k) \\ &= \frac{(n-k)!}{n!} \mathbb{E} \left[\sum_{\sigma \in \mathfrak{S}(k, n)} \mathbb{1}_{r(\sigma(1), T_{n+1}) = i_1, \dots, r(\sigma(k), T_{n+1}) = i_k} \right] \\ &= \frac{(n-k)!}{n!} \mathbb{E} \left[\prod_{s=1}^r (\#\{1 \leq j \leq n : r(j, T_{n+1}) = j_s\})_{m_s} \right] \\ &= \frac{(n-k)!}{n!} \mathbb{E} \left[\prod_{s=1}^r (X_{j_s})_{m_s} \right]. \end{aligned}$$

The first equality comes from the fact that any permutation of a parking function is still a parking function, thus any permutation induces a bijection in P_n . The second equality comes from Theorem III.B.1 and the last from Corollary III.B.2. This completes the proof. 

C Convergence for the total variation distance

In this section we suppose that $k_n = o(\sqrt{n})$. We will write k instead of k_n to make notation lighter but keep in mind that k depends on n . The goal of this section is to show item (i) of Theorem III.A.1.

C.i Probability that the parking places are distinct

The first step is to reduce the problem to distinct parking places, in this case Equation (B.2) becomes easier. To this end we introduce the set of distinct indices $D_n := \{(u_1, \dots, u_k) \in \llbracket 1, n \rrbracket^k : i \neq j \Rightarrow u_i \neq u_j\}$. We also introduce the set $G_n := \{(u_1, \dots, u_k) \in \llbracket 1, n \rrbracket^k : \mathbb{P}(\pi_n(1) = u_1, \dots, \pi_n(k) = u_k) \geq (n-k)!/n!\}$ and the quantity

$$\delta(k, n) := \sum_{\substack{(i_1, \dots, i_k) \\ \in D_n \cap G_n}} \left[\mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) - \frac{(n-k)!}{n!} \right]. \quad (\text{C.1})$$

The next Lemma shows that the first k parking places of a uniform parking function are all distinct with high probability. It also shows that if $\delta(k, n)$ is a $O(k/\sqrt{n})$ then so is $d_{TV}(k, n)$. Recall that $(U_n(i))_{1 \leq i \leq n}$ are i.i.d. uniformly distributed in $\llbracket 1, n \rrbracket$.

Lemme III.C.1

We have

- (i) $\mathbb{P}((U_n(1), \dots, U_n(k)) \in D_n) = 1 + O\left(\frac{k}{\sqrt{n}}\right)$,
- (ii) $\mathbb{P}((\pi_n(1), \dots, \pi_n(k)) \in D_n) = 1 + O\left(\frac{k}{\sqrt{n}}\right)$,
- (iii) $\delta(k, n) = O\left(\frac{k}{\sqrt{n}}\right) \implies d_{TV}(k, n) = O\left(\frac{k}{\sqrt{n}}\right)$.

Preuve

Let μ_n be the law of $(\pi_n(1), \dots, \pi_n(k))$ and ν_n the law of $(U_n(1), \dots, U_n(k))$, with support on the same finite space $E_n := \llbracket 1, n \rrbracket^k$. First we check (i). By Markov's inequality

$$\nu_n(D_n^c) = \mathbb{P}\left(\sum_{r < s} \mathbb{1}_{U_n(r) = U_n(s)} \geq 1\right) \leq \sum_{r < s} \mathbb{P}(U_n(r) = U_n(s)) = \sum_{r < s} \frac{1}{n} = \frac{1}{n} \frac{k(k-1)}{2}.$$

Since $k = o(\sqrt{n})$ we have that

$$\frac{1}{n} \frac{k(k-1)}{2} = O\left(\frac{k^2}{n}\right) = O\left(\frac{k}{\sqrt{n}}\right).$$

Now we check (ii). To do so, we will use the Prüfer encoding (or a slight variant thereof) of a rooted Cayley tree $t \in \mathfrak{C}_n$ into a sequence $(p_1, \dots, p_{n-2}) \in \llbracket 0, n-1 \rrbracket^{n-2}$, which we now explain. For $a \in \llbracket 1, n-1 \rrbracket$, define $p(t, a)$ to be the label of the parent of the vertex labeled a in t . Also, define $\ell(t)$ as the biggest leaf label of t , and t^* the tree t obtained after removing the leaf labeled $\ell(t)$ and its adjacent edge. Finally we define the sequence of trees, $t_1 := t$, $t_i := t_{i-1}^*$ for $2 \leq i \leq n-2$. The Prüfer encoding of t is then defined as $p_i := p(t_i, \ell(t_i))$. For example, the Prüfer encoding of the tree in figure B.1 is $(8, 8, 6, 0, 3, 0, 3, 3)$. The key property of this encoding is that it is a bijection between the sets \mathfrak{C}_n and $\llbracket 0, n-1 \rrbracket^{n-2}$. Now, let T_{n+1} be a uniform Cayley tree in \mathfrak{C}_{n+1} . Theorem III.B.1 implies that $\mu_n(D_n)$ is equal to the probability that the vertices labeled 1 to k in T_{n+1} have distinct parents. Let (v_1, \dots, v_k) be a random vector with uniform

distribution in D_n independent of T_{n+1} . Since the distribution of T_{n+1} is invariant under permutation of the labels, the previous probability is also equal to the probability that the vertices labeled v_1, \dots, v_k have distinct parents in T_{n+1} . Let (p_1, \dots, p_{n-1}) be the Prüfer encoding of the tree T_{n+1} . We complete this vector with $p_n := 0$ (this comes from the fact that t_{n-2} has two vertices, one of them being the root labeled 0). Since T_{n+1} is uniformly distributed in \mathcal{C}_{n+1} , the vector (p_1, \dots, p_{n-1}) is uniformly distributed in $\llbracket 0, n \rrbracket^{n-1}$. From the previous discussion and the definition of the Prüfer encoding we deduce that

$$\mu_n(D_n) = \mathbb{P}((p_{v_1}, \dots, p_{v_k}) \in D_n).$$

Consider the event $Z_n := \{v_1, \dots, v_k \neq n\}$. Under this event, it is easy to see that $(p_{v_1}, \dots, p_{v_k})$ has the same law as k i.i.d random variables uniformly distributed in $\llbracket 0, n \rrbracket$. So from (i) we have

$$\mu_n(D_n) \geq \mathbb{P}((p_{v_1}, \dots, p_{v_k}) \in D_n | Z_n) \mathbb{P}(Z_n) = \left(1 + O\left(\frac{k}{\sqrt{n}}\right)\right) \mathbb{P}(Z_n).$$

To conclude, notice that $\mathbb{P}(Z_n) = 1 - k/n$.

Finally we show (iii). Assume that $\delta(k, n) = O(k/\sqrt{n})$. For all i_1, \dots, i_k denote by Δ_{i_1, \dots, i_k} the quantity $(\mathbb{P}(\pi_n(1) = i_1, \dots, \pi_n(k) = i_k) - (n-k)!/n!)$. Notice that $n^k(n-k)!/n! - 1 = O(k/\sqrt{n})$ so

$$2d_{TV}(k, n) = \sum_{i_1, \dots, i_k=1}^n |\Delta_{i_1, \dots, i_k}| + O\left(\frac{k}{\sqrt{n}}\right).$$

Denote by d_n^+ the sum of Δ_{i_1, \dots, i_k} over the indices in G_n and d_n^- the opposite of the sum over the indices in $E_n \setminus G_n$. We have that


$$d_n^+ - d_n^- = \sum_{i_1, \dots, i_k=1}^n \Delta_{i_1, \dots, i_k} = 1 - \frac{n^k(n-k)!}{n!} = O\left(\frac{k}{\sqrt{n}}\right).$$

The last two equalities imply that

$$2d_{TV}(k, n) = d_n^+ + d_n^- + O\left(\frac{k}{\sqrt{n}}\right) = 2d_n^+ + O\left(\frac{k}{\sqrt{n}}\right).$$

In conclusion we just need to show that d_n^+ is a $O(k/\sqrt{n})$. Notice that

$$d_n^+ = \delta(k, n) + \mu_n(D_n^c \cap G_n) - \nu_n(D_n^c \cap G_n) \times \frac{n^k(n-k)!}{n!}.$$

From (i), (ii) and the assumption on $\delta(k, n)$ we deduce that d_n^+ is indeed a $O(k/\sqrt{n})$. 

To prove (i) of Theorem III.A.1 it remains to show that $\delta(k, n) = O(k/\sqrt{n})$. This is the goal of the next three sections.

C.ii A monotonicity argument

In this section we bound the terms $\mathbb{E}_n[X_{i_1} \cdots X_{i_k}]$ that appear in (B.2) when i_1, \dots, i_k are distinct with terms involving $\mathbb{E}_n[S_{i_1} \cdots S_{i_k}]$ since the latter are more manageable. More precisely, the aim of this section is to prove the following result.

Proposition III.C.2

Fix $i_1, \dots, i_k \in \llbracket 1, n \rrbracket$ pairwise distinct. We have

$$i_1 \cdots i_k \mathbb{E}_n[X_{i_1} \cdots X_{i_k}] \leq \mathbb{E}_n[(S_{i_1} + i_1) \cdots (S_{i_k} + i_k)] \quad (\text{C.2})$$

To prove Proposition III.C.2 we first state a really useful lemma which, put in simple terms, says that the steps of the random walk S tend to decrease under \mathbb{P}_n .

Lemme III.C.3

Fix $n \geq k \geq 1$ and $m_1, \dots, m_k \geq 1$. Let $1 \leq i_1 < \dots < i_k \leq n$ and $1 \leq j_1 < \dots < j_k \leq n$ such that $j_r \leq i_r$ for all $1 \leq r \leq k$. Let $f : \mathbb{N} \times \mathbb{N} \setminus \{0\} \mapsto [0, \infty)$ be a nonnegative function such that $f(0, m) = 0$, then

$$\mathbb{E}_n[f(X_{i_1}, m_1) \cdots f(X_{i_k}, m_k)] \leq \mathbb{E}_n[f(X_{j_1}, m_1) \cdots f(X_{j_k}, m_k)]. \quad (\text{C.3})$$


Preuve du Lemme III.C.3

To make notations lighter, we define the random vector $\mathbf{X} := (X_1, \dots, X_{n+1})$ and write \mathbf{x} as a shortcut to designate an element (x_1, \dots, x_{n+1}) of \mathbb{N}^{n+1} . Let $s := \min\{r \geq 1 : j_r < i_r\}$. We only need to treat the case where $i_r = j_r$ for all $r \neq s$ and $j := j_s = i_s - 1$ (the general result can then be obtained by induction). Let $\sigma = (j \ j+1) \in \mathfrak{S}_{n+1}$ be the permutation that transposes j and $j+1$. Let $\mathcal{E}_n := \{\mathbf{x} \in \mathbb{N}^{n+1} : (x_1 - 1) + \dots + (x_t - 1) = -1 \text{ iff } t = n+1\}$ and let $\mathcal{E}'_n := \{\mathbf{x} \in \mathcal{E}_n : x_{j+1} > 0 \text{ or } (x_1 - 1) + \dots + (x_{j-1} - 1) > 0\}$. Notice that $(x_1, \dots, x_{n+1}) \mapsto (x_{\sigma(1)}, \dots, x_{\sigma(n+1)})$ is a bijection on \mathcal{E}'_n .

$$\begin{aligned} \mathbb{E} \left[\prod_{r=1}^k f(X_{i_r}, m_r) \mathbb{1}_{\tau_{-1}=n+1} \right] &= \sum_{\mathbf{x} \in \mathcal{E}_n} \prod_{r=1}^k f(x_{i_r}, m_r) \mathbb{P}[\mathbf{X} = \mathbf{x}] \\ &= \sum_{\mathbf{x} \in \mathcal{E}'_n} \prod_{r=1}^k f(x_{\sigma(i_r)}, m_r) \mathbb{P}[\mathbf{X} = \mathbf{x}] \\ &\quad + \sum_{\mathbf{x} \in \mathcal{E}_n \setminus \mathcal{E}'_n} \prod_{r=1}^k f(x_{i_r}, m_r) \mathbb{P}[\mathbf{X} = \mathbf{x}]. \end{aligned}$$

Notice that if $\mathbf{x} \in \mathcal{E}_n \setminus \mathcal{E}'_n$ then $f(x_{j+1}, m_{j+1}) = f(0, m_{j+1}) = 0$; in particular, since f is nonnegative, $f(x_{i_1}, m_1) \cdots f(x_{i_k}, m_k) \leq f(x_{\sigma(i_1)}, m_1) \cdots f(x_{\sigma(i_k)}, m_k)$. Finally

$$\begin{aligned} \mathbb{E} \left[\prod_{r=1}^k f(X_{i_r}, m_r) \mathbb{1}_{\tau_{-1}=n+1} \right] &\leq \sum_{\mathbf{x} \in \mathcal{E}_n} \prod_{r=1}^k f(x_{\sigma(i_r)}, m_r) \mathbb{P}[\mathbf{X} = \mathbf{x}] \\ &= \mathbb{E} \left[\prod_{r=1}^k f(X_{\sigma(i_r)}, m_r) \mathbb{1}_{\tau_{-1}=n+1} \right] \\ &= \mathbb{E} \left[\prod_{r=1}^k f(X_{j_r}, m_r) \mathbb{1}_{\tau_{-1}=n+1} \right] \end{aligned}$$

which is enough to conclude. 

Remarque

In Lemma III.C.3 we can for instance take $f(x, m) = x^m$ or $f(x, m) = (x)_m$. Notice that in Lemma III.C.3 the indices $(i_r)_r$ must be pairwise distinct as well as the indices $(j_r)_r$. In the proof of Proposition III.C.2 we extend the result, when $f(x, m) = x^m$, to the case where only the $(i_r)_r$ are pairwise distinct.

Preuve de la Proposition III.C.2

First we show the following inequality. Fix $n \geq k \geq 1$. Let $1 < i_1 < \dots < i_k \leq n$, $1 \leq j_1 \leq \dots \leq j_k \leq n$ be such that $j_r \leq i_r$ for all $1 \leq r \leq k$. Then

$$\mathbb{E}_n[X_{i_1} \cdots X_{i_k}] \leq \mathbb{E}_n[X_{j_1} \cdots X_{j_k}]. \quad (\text{C.4})$$

To show (C.4) it is actually enough to show the following result : let $J \subset \llbracket 1, n \rrbracket$ and $2 \leq i \leq n$ such that i and $i-1$ do not belong to J . Let $m_j \geq 1$ for $j \in J$ and $m \geq 1$. Then


$$\mathbb{E}_n \left[X_{i-1}^m X_i \prod_{j \in J} X_j^{m_j} \right] \leq \mathbb{E}_n \left[X_{i-1}^{m+1} \prod_{j \in J} X_j^{m_j} \right]. \quad (\text{C.5})$$

Inequality (C.4) can then be obtained by induction using Lemma III.C.3 and (C.5). By Young's inequality :

$$X_{i-1}^m X_i \leq \frac{m}{m+1} X_{i-1}^{m+1} + \frac{1}{m+1} X_i^{m+1}.$$

Combining this with Lemma III.C.3 gives (C.5) and concludes the proof of (C.4). Now, using inequality (C.4) we obtain

$$i_1 \cdots i_k \mathbb{E}_n[X_{i_1} \cdots X_{i_k}] \leq \sum_{\forall r, j_r \leq i_r} \mathbb{E}_n[X_{j_1} \cdots X_{j_k}] = \mathbb{E}_n[(S_{i_1} + i_1) \cdots (S_{i_k} + i_k)]$$

which concludes the proof of Proposition III.C.2. 

C.iii Bounding the moments of a random walk conditioned to be an excursion

The goal of this section is to find bounds for the moments of the random walk S conditioned to be an excursion. More precisely, the aim of this section is to show the following result.

Proposition III.C.4

There exists a constant $C > 0$ such that for all $n \geq 2$, $1 \leq k \leq n-1$ and $d \geq 1$

(i) we have

$$\mathbb{E}_n[S_k^d] \leq (Cdn)^{d/2}, \quad (\text{C.6})$$

(ii) and

$$\mathbb{E}_n[S_{n-k}^d] \leq \left(\frac{n}{n-k}\right)^{3/2} (Cdk)^{d/2}, \quad (\text{C.7})$$

(iii) as well as

$$\mathbb{E}_n[S_k^d] \leq \left(\frac{n}{n-k}\right)^{3/2} (Cd\sqrt{k})^d. \quad (\text{C.8})$$

Remarque

Proposition III.C.4 (i) and (ii) actually hold true for any random walk $(S_n)_{n \geq 0}$ starting from 0 with i.i.d. increments all having the distribution μ whose support is $\mathbb{N} \cup \{-1\}$ and such that μ has mean 0 and finite variance. However, (iii) uses in addition the fact that a Poisson random walk has a sub-exponential tail (see e.g. [100, Example 3]), namely, for all $k \geq 1$ and $x \geq 0$:

$$\mathbb{P}(S_k \geq x) \leq \exp\left(-\frac{x^2}{2(k+x)}\right). \tag{C.9}$$

To prove Proposition III.C.4 we will use the following lemma whose proof is postponed at the end of this section. This lemma is similar to the cyclic lemma in spirit but instead of conditioning the walk to be an excursion we only condition it to stay positive.

Lemme III.C.5

Let $n \geq 1$ and $F : \mathbb{R}^n \rightarrow [0, +\infty)$ be invariant under cyclic shifts. Then

$$\mathbb{E}\left[F(X_1, \dots, X_n) \mathbb{1}_{S_1, \dots, S_n > 0}\right] \leq \frac{1}{n} \mathbb{E}\left[F(X_1, \dots, X_n) (S_n \wedge n) \mathbb{1}_{S_n > 0}\right]. \tag{C.10}$$

Preuve de la Proposition III.C.4

We recall that C denotes a constant which may vary line to line. For (i), according to [3, Eq. (32)] the maximum of the excursion of S has a sub-Gaussian tail, namely there exist constants $C, \alpha > 0$ such that for all $n \geq 1$ and $x \geq 0$:

$$\mathbb{P}_n(M_n \geq x) \leq C e^{-\alpha x^2/n}$$

where $M_n := \max\{S_0, \dots, S_n\}$ is the maximum of the walk S on $[0, n]$. So we have

$$\begin{aligned} \mathbb{E}_n\left[n^{-d/2} S_k^d\right] &\leq \mathbb{E}_n\left[n^{-d/2} M_n^d\right] = \int_0^\infty dx x^{d-1} \mathbb{P}_n(M_n \geq \sqrt{n}x) \\ &\leq \int_0^\infty C dx x^{d-1} e^{-\alpha x^2} \leq C^d d^{d/2}. \end{aligned}$$

This shows (i).

The following computation is a common starting point to show both (ii) and (iii). Let $H(S_k, x)$ be either the indicator function $\mathbb{1}_{S_k=x}$ or $\mathbb{1}_{S_k \geq x}$ with $x > 0$. Using the fact that $\mathbb{P}(\tau_{-1} = n+1)$ is equivalent to a constant times $n^{-3/2}$ (see e.g. [78, Eq. (10)]), then the Markov property and finally the cyclic lemma (see e.g. [88, Section 6.1]) we get

$$\begin{aligned} \mathbb{E}_n[H(S_k, x)] &\leq C n^{3/2} \mathbb{E}\left[H(S_k, x) \mathbb{1}_{\tau_{-1}=n+1}\right] \\ &\leq C n^{3/2} \mathbb{E}\left[H(S_k, x) \mathbb{1}_{S_1, \dots, S_k \geq 0} \mathbb{P}(\tau_{-1-S_k} = n-k)\right] \\ &\leq C n^{3/2} \mathbb{E}\left[H(S_k, x) \mathbb{1}_{S_1, \dots, S_k \geq 0} \frac{1+S_k}{n-k} \mathbb{P}(S'_{n-k} = -1 - S_k)\right] \end{aligned}$$

where S' is independent of S and has the same distribution. Now we use Janson's inequality [67, Lemma

2.1] which states that $\mathbb{P}(S_r = -m) \leq Cr^{-1/2} e^{-\lambda m^2/r}$ for all $r \geq 1$ and $m \geq 0$ to get

$$\begin{aligned} \mathbb{E}_n[H(S_k, x)] &\leq Cn^{3/2} \mathbb{E} \left[H(S_k, x) \mathbb{1}_{S_1, \dots, S_k \geq 0} \frac{1 + S_k}{(n-k)^{3/2}} e^{-\lambda \frac{(1+S_k)^2}{n-k}} \right] \\ &\leq C \left(\frac{n}{n-k} \right)^{3/2} \mathbb{E} \left[H(S_k, x) \mathbb{1}_{S_1, \dots, S_k \geq 0} S_k e^{-\lambda \frac{S_k^2}{n-k}} \right]. \end{aligned} \quad (\text{C.11})$$

To prove (ii) we first define $T_k := \max\{0 \leq i \leq k : S_i = 0\}$. Denote by $p_k(x)$ the probability of the event $\{S_k = x, S_1, \dots, S_k \geq 0\}$. Using the Markov property, we obtain

$$\begin{aligned} p_k(x) &= \sum_{i=0}^{k-1} \mathbb{P}(S_k = x, S_1, \dots, S_k \geq 0, T_k = i) \\ &= \sum_{i=0}^{k-1} \mathbb{P}(S_1, \dots, S_i \geq 0, S_i = 0) \mathbb{P}(S_{k-i} = x, S_1, \dots, S_{k-i} > 0). \end{aligned}$$

We apply Lemma III.C.5 and the local limit theorem (see e.g. [62, Theorem 4.2.1]), which gives a constant $C > 0$ such that $\mathbb{P}(S_{k-i} = x) \leq C(k-i)^{-1/2}$ for every $k > i$ and $x \in \mathbb{Z}$,

$$\begin{aligned} p_k(x) &\leq \sum_{i=0}^{k-1} \mathbb{P}(S_1, \dots, S_i \geq 0, S_i = 0) \frac{x}{k-i} \mathbb{P}(S_{k-i} = x) \\ &\leq Cx \sum_{i=0}^{k-1} \mathbb{P}(S_1, \dots, S_i \geq 0, S_i = 0) \frac{1}{(k-i)^{3/2}}. \end{aligned}$$

Notice that

$$\mathbb{P}(S_1, \dots, S_i \geq 0, S_i = 0) = e \mathbb{P}(\tau_{-1} = i+1) \leq \frac{C}{(i+1)^{3/2}}.$$

So, finally we have

$$p_k(x) \leq Cx \sum_{i=0}^{k-1} \frac{1}{((i+1)(k-i))^{3/2}} = \frac{Cx}{(k+1)^{3/2}} \sum_{i=1}^{k-1} \left(\frac{1}{i+1} + \frac{1}{k-i} \right)^{3/2} \leq \frac{Cx}{k^{3/2}}.$$

Putting the previous inequality in (C.11) with $H(S_k, x) = \mathbb{1}_{S_k=x}$ and replacing k by $n-k$ gives

$$\mathbb{P}_n(S_{n-k} = x) \leq C \left[\frac{n}{k(n-k)} \right]^{3/2} x^2 e^{-\lambda \frac{x^2}{k}}.$$

We can now bound the d -th moment of S_{n-k}

$$\begin{aligned} \mathbb{E}_n[S_{n-k}^d] &\leq C \left[\frac{n}{k(n-k)} \right]^{3/2} \int_0^\infty x^{d+2} e^{-\lambda \frac{x^2}{k}} dx \leq C^d k^{d/2} \left[\frac{n}{n-k} \right]^{3/2} \int_0^\infty x^{d+2} e^{-x^2} dx \\ &\leq C^d k^{d/2} \left[\frac{n}{n-k} \right]^{3/2} d^{d/2}. \end{aligned}$$

This concludes the proof of (ii).

To prove (iii) we follow the same principle. Using the Markov property and Lemma III.C.5 we have

$$\begin{aligned} \mathbb{E} \left[\mathbb{1}_{S_k \geq x} \mathbb{1}_{S_1, \dots, S_k \geq 0} S_k \right] &= \sum_{i=0}^{k-1} \mathbb{E} \left[\mathbb{1}_{S_k \geq x} \mathbb{1}_{S_1, \dots, S_k \geq 0} S_k \mathbb{1}_{T_k=i} \right] \\ &= \sum_{i=0}^{k-1} \mathbb{P}(S_1, \dots, S_i \geq 0, S_i = 0) \mathbb{E} \left[\mathbb{1}_{S_{k-i} \geq x} \mathbb{1}_{S_1, \dots, S_{k-i} > 0} S_{k-i} \right] \\ &\leq \sum_{i=0}^{k-1} \mathbb{P}(S_1, \dots, S_i \geq 0, S_i = 0) \frac{1}{k-i} \mathbb{E} \left[\mathbb{1}_{S_{k-i} \geq x} S_{k-i}^2 \right]. \end{aligned}$$

Then, we apply the Cauchy-Schwarz inequality

$$\mathbb{E}[\mathbb{1}_{S_{k-i} \geq x} S_{k-i}^2] \leq \mathbb{P}(S_{k-i} \geq x)^{1/2} \mathbb{E}[S_{k-i}^4]^{1/2} \leq C(k-i) \mathbb{P}(S_{k-i} \geq x)^{1/2}.$$

The last inequality comes from an explicit computation of the fourth central moment of a Poisson distribution. We combine the last inequality with (C.9) to get

$$\mathbb{E}[\mathbb{1}_{S_{k-i} \geq x} S_{k-i}^2] \leq C(k-i) e^{-\frac{x^2}{4(k-i+x)}} \leq C(k-i) e^{-\frac{x^2}{4(k+x)}}.$$

Putting everything together we obtain


$$\begin{aligned} \mathbb{E}[\mathbb{1}_{S_k \geq x} \mathbb{1}_{S_1, \dots, S_k \geq 0} S_k] &\leq C e^{-\frac{x^2}{4(k+x)}} \sum_{i=0}^{k-1} \mathbb{P}(S_1, \dots, S_i \geq 0, S_i = 0) \\ &\leq C e^{-\frac{x^2}{4(k+x)}} \sum_{i=0}^{k-1} e \mathbb{P}(\tau_{-1} = i+1) \leq C e^{-\frac{x^2}{4(k+x)}}. \end{aligned}$$

We combine the last inequality with (C.11) to get

$$\mathbb{E}_n[k^{-d/2} S_k^d] = \int_0^\infty dx x^{d-1} \mathbb{P}_n(S_k \geq \sqrt{k}x) dx \leq C \left[\frac{n}{n-k} \right]^{3/2} \int_0^\infty dx x^{d-1} e^{-\frac{kx^2}{4(k+x\sqrt{k})}} dx.$$

We cut the last integral into two parts and obtain

$$\int_0^\infty dx x^{d-1} e^{-\frac{kx^2}{4(k+x\sqrt{k})}} dx \leq \int_0^{\sqrt{k}} dx x^{d-1} e^{-\frac{x^2}{8}} dx + \int_{\sqrt{k}}^\infty dx x^{d-1} e^{-\frac{x}{8}} dx.$$


Noticing that the last two integrals are both smaller than $C^d d^d$, this concludes the proof of (iii). 

We now prove Lemma III.C.5

Preuve du Lemme III.C.5

For a vector $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $i \in \llbracket 0, n-1 \rrbracket$ denote by \mathbf{x}^i the i -th cyclic permutation of \mathbf{x} , namely $\mathbf{x}^i := (x_{1+i}, \dots, x_n, x_1, \dots, x_i)$. Consider the set $A_n := \{(x_1, \dots, x_n) \in \mathbb{N}^n : \forall k \in \llbracket 1, n \rrbracket, \sum_{i=1}^k (x_i - 1) > 0\}$ and write $\mathbf{X} := (X_1, \dots, X_n)$. Then

$$\begin{aligned} \mathbb{E}[F(X_1, \dots, X_n) \mathbb{1}_{S_1, \dots, S_n > 0}] &= \mathbb{E}[F(\mathbf{X}) \mathbb{1}_{\mathbf{X} \in A_n}] = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[F(\mathbf{X}^i) \mathbb{1}_{\mathbf{X}^i \in A_n}] \\ &= \frac{1}{n} \mathbb{E}\left[F(\mathbf{X}) \sum_{i=0}^{n-1} \mathbb{1}_{\mathbf{X}^i \in A_n}\right] \\ &\leq \frac{1}{n} \mathbb{E}[F(\mathbf{X})(S_n \wedge n) \mathbb{1}_{S_n > 0}]. \end{aligned}$$

The inequality comes from the fact that the number of cyclic shifts of \mathbf{X} such that $\mathbf{X}^i \in A_n$ is almost surely bounded by $(S_n \wedge n) \mathbb{1}_{S_n > 0}$. 

C.iv Proof of Theorem III.A.1 (i)

Recall we want to show that $d_{TV}(k, n) = O(k/\sqrt{n})$. In section C.i we have shown that it is enough to show $\delta(k, n) = O(k/\sqrt{n})$ where the definition of $\delta(k, n)$ is given by (C.1). Thanks to Proposition III.B.3 this quantity can be rewritten as

$$\delta(k, n) = \frac{n^k(n-k)!}{n!} \int_{\Lambda_k} \left[\mathbb{E}_n [X_{\lceil nt_1 \rceil} \cdots X_{\lceil nt_k \rceil}] - 1 \right] dt_1 \cdots dt_k$$

where $\Lambda_k := \{(t_1, \dots, t_k) \in (0, 1]^k : (\lceil nt_1 \rceil, \dots, \lceil nt_k \rceil) \in D_n \cap G_n\}$. As was already mentioned in section C.i, we have the asymptotic $n^k(n-k)!/n! = 1 + O(k/\sqrt{n})$ so, for our purpose, it is sufficient to bound the integral

$$I(k, n) := \int_{\Lambda_k} \left[\mathbb{E}_n [X_{\lceil nt_1 \rceil} \cdots X_{\lceil nt_k \rceil}] - 1 \right] dt_1 \cdots dt_k.$$

Using inequality (C.2) we obtain

$$I(k, n) \leq \int_{[0,1]^k} \left[\mathbb{E}_n \left[\left(\frac{S_{\lceil nt_1 \rceil}}{\lceil nt_1 \rceil} + 1 \right) \cdots \left(\frac{S_{\lceil nt_k \rceil}}{\lceil nt_k \rceil} + 1 \right) \right] - 1 \right] dt_1 \cdots dt_k. \quad (\text{C.12})$$

Notice that Proposition III.C.4 (i) and (iii) imply that for all $0 \leq k \leq n$

$$\mathbb{E}_n [S_k^d] \leq (Cd\sqrt{k})^d. \quad (\text{C.13})$$

Indeed, Proposition III.C.4 (i) covers the case $k \geq n/2$ of equation (C.13) and (iii) covers the case $k < n/2$ (notice that the cases $k = 0$ and $k = n$ are trivial, since, conditionally on $\tau_{-1} = n + 1$, $S_0 = S_n = 0$). Hölder's inequality shows that for $0 \leq i_1, \dots, i_d \leq n$

$$\mathbb{E}_n [S_{i_1} \cdots S_{i_d}] \leq (Cd)^d \sqrt{i_1 \cdots i_d}. \quad (\text{C.14})$$

Expanding the products in (C.12) and using (C.14) gives

$$I(k, n) \leq \sum_{d=1}^k \binom{k}{d} (Cd)^d n^{-d/2} \int_{[0,1]^d} \frac{dt_1 \cdots dt_d}{(t_1 \cdots t_d)^{1/2}}.$$

Notice that $t \mapsto t^{-1/2}$ is integrable so

$$I(k, n) \leq \sum_{d=1}^k \binom{k}{d} (Cd)^d n^{-d/2}.$$

Using the bound $\binom{k}{d} \leq (ke/d)^d$,

$$I(k, n) \leq \sum_{d=1}^k \left(Ce \frac{k}{\sqrt{n}} \right)^d.$$

Since $k = o(\sqrt{n})$ we conclude that $I(k, n) = O(k/\sqrt{n})$.

D Convergence for the Kolmogorov distance

In this section we suppose that $k_n = o(n)$ and $\sqrt{n} = o(k_n)$. We will write k instead of k_n to make notation lighter but keep in mind that k depends on n . The goal of this section is to show Theorem III.A.1 (ii). The following lemma allows us to replace the cumulative probability in (A.2) with the term $\mathbb{E}_n[(S_{i_1} + i_1) \cdots (S_{i_n} + i_n)]$ which is more manageable.

Lemme III.D.1

There is a constant $C > 0$ such that

$$d_K(k, n) \leq \frac{1}{n^k} \max_{1 \leq i_1, \dots, i_k \leq n} \left| \mathbb{E}_n[(S_{i_1} + i_1) \cdots (S_{i_k} + i_k)] - i_1 \cdots i_k \right| + \frac{Ck}{n}. \quad (\text{D.1})$$

Before proving this lemma we show how it implies Theorem III.A.1 (ii). We will also need the following simple lemma which extends [24, Eq. (27.5)]

Lemme III.D.2

Let $r \geq 1$, w_1, \dots, w_r and z_1, \dots, z_r be complex numbers of modulus smaller or equal to, respectively, $a > 0$ and $b > 0$. Then

$$\left| \prod_{i=1}^r w_i - \prod_{i=1}^r z_i \right| \leq \sum_{i=1}^r |w_i - z_i| a^{r-i} b^{i-1}. \quad (\text{D.2})$$

Preuve du Lemme III.D.2

The result readily follows from the identity

$$\prod_{i=1}^r w_i - \prod_{i=1}^r z_i = (w_1 - z_1) \prod_{i=2}^r w_i + z_1 \left(\prod_{i=2}^r w_i - \prod_{i=2}^r z_i \right).$$


Preuve du Théorème III.A.1 (ii)

Let $1/2 < \alpha < 1$. We define a sequence of intervals I_1, \dots, I_{M+3} (depending on n) in the following way.

$$I_1 := [1, n-k), \quad I_2 := [n-k, n-k^\alpha), \quad I_3 := [n-k^\alpha, n-k^{\alpha^2}), \quad \dots \\ \dots, \quad I_{M+1} := [n-k^{\alpha^{M-1}}, n-k^{\alpha^M}), \quad I_{M+2} := [n-k^{\alpha^M}, n-n/k), \quad I_{M+3} := [n-n/k, n]$$

where M is the biggest integer such that $n-k^{\alpha^M} \leq n-n/k$. Let $1 \leq i_1, \dots, i_k \leq n$, using Lemma III.D.2 (with $a = b = 1$ and noticing that $(S_i + i)/n$ under \mathbb{P}_n is almost surely smaller than 1 for every i) we decompose the quantity $\mathbb{E}_n[(S_{i_1} + i_1) \cdots (S_{i_k} + i_k) - i_1 \cdots i_k]$ depending on what intervals the i_j 's belong to.

$$\mathbb{E}_n \left[\left(\frac{S_{i_1} + i_1}{n} \right) \cdots \left(\frac{S_{i_k} + i_k}{n} \right) - \frac{i_1}{n} \cdots \frac{i_k}{n} \right] \leq \sum_{m=1}^{M+3} \mathbb{E}_n \left[\prod_{i_j \in I_m} \frac{S_{i_j} + i_j}{n} - \prod_{i_j \in I_m} \frac{i_j}{n} \right]. \quad (\text{D.3})$$

Fix $m \in \{1, \dots, M+3\}$ and denote by i_1, \dots, i_{r_m} the i_j 's that belong to I_m . If $m = 1$ then by Lemma III.D.2 and Proposition III.C.4 (i)

$$\mathbb{E}_n \left[\prod_{i_j \in I_1} \frac{S_{i_j} + i_j}{n} - \prod_{i_j \in I_1} \frac{i_j}{n} \right] \leq \frac{1}{n} \sum_{j=1}^{r_1} \mathbb{E}_n[S_{i_j}] \left(\frac{n-k}{n} \right)^{j-1} \leq \frac{C\sqrt{n}}{n} \frac{n}{k} = \frac{C\sqrt{n}}{k}.$$

If $2 \leq m \leq M+2$ we follow the same principle but we use Proposition III.C.4 (ii) instead

$$\mathbb{E}_n \left[\prod_{i_j \in I_m} \frac{S_{i_j} + i_j}{n} - \prod_{i_j \in I_m} \frac{i_j}{n} \right] \leq \frac{1}{n} \sum_{j=1}^{r_m} \mathbb{E}_n[S_{i_j}] \left(\frac{n-k^{\alpha^{m-1}}}{n} \right)^{j-1} \leq \frac{Ck^{\alpha^{m-2}/2}}{n} \frac{n}{k^{\alpha^{m-1}}} = \frac{C}{k^{(\alpha-1/2)\alpha^{m-2}}}.$$



The previous computation works in the case $m = M + 2$ because the maximality of M implies that $n - n/k \leq n - k^{\alpha^{M+1}}$. Finally if $m = M + 3$


$$\mathbb{E}_n \left[\prod_{i_j \in I_{M+3}} \frac{S_{i_j} + i_j}{n} - \prod_{i_j \in I_{M+3}} \frac{i_j}{n} \right] \leq \frac{1}{n} \sum_{j=1}^{r_{M+3}} \mathbb{E}_n[S_{i_j}] \leq r_{M+3} \frac{C}{n} \left(\frac{n}{k}\right)^{1/2} \leq C \left(\frac{k}{n}\right)^{1/2}.$$

Notice that $k^{\alpha^M} \geq n/k$ so for all $0 \leq m \leq M$, $k^{\alpha^{M-m}} \geq (n/k)^{\alpha^{-m}}$. Summing the preceding bounds for $2 \leq m \leq M + 2$ gives

$$\begin{aligned} \sum_{m=2}^{M+2} \mathbb{E}_n \left[\prod_{i_j \in I_m} \frac{S_{i_j} + i_j}{n} - \prod_{i_j \in I_m} \frac{i_j}{n} \right] &\leq C \sum_{m=0}^M \frac{1}{k^{(\alpha-1/2)\alpha^m}} = C \sum_{m=0}^M \frac{1}{k^{(\alpha-1/2)\alpha^{M-m}}} \\ &\leq C \sum_{m=0}^M \left(\frac{k}{n}\right)^{(\alpha-1/2)\alpha^{-m}} \\ &\leq C \left(\frac{k}{n}\right)^{(\alpha-1/2)} + C \sum_{m=1}^M \left(\frac{k}{n}\right)^{-(\alpha-1/2)e \ln(\alpha)m} \\ &\leq C \left(\frac{k}{n}\right)^{(\alpha-1/2)} + C \left(\frac{k}{n}\right)^{-(\alpha-1/2)e \ln(\alpha)} \end{aligned}$$

where we used the fact that $-e \ln(\alpha)m \leq 1/\alpha^m$ for all $m \geq 1$. If we take α such that $\alpha - 1/2 = \exp(-e^{-1}) - 1/2 \approx 0.1922$ then the last quantity is a $O(k/n)^{0.19}$. Putting everything together we finally obtain that

$$\frac{1}{n^k} \max_{1 \leq i_1, \dots, i_k \leq n} \left| \mathbb{E}_n \left[\prod_{r=1}^k (S_{i_r} + i_r) \right] - \prod_{r=1}^k i_r \right| \leq C \left[\frac{\sqrt{n}}{k} + \left(\frac{k}{n}\right)^{0.19} + \left(\frac{k}{n}\right)^{1/2} \right].$$

Combining the last display with Lemma III.D.1 gives the desired result. 

Now we prove Lemma III.D.1.

Preuve du Lemme III.D.1

Define the empirical distribution function of π_n , namely, for all $i \in \{1, \dots, n\}$:

$$F_n(i) := \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\pi_n(j) \leq i}.$$

As suggested in [45] we can use a result of Bobkov [25, Theorem 1.1]. The sequence $(\pi_n(1), \dots, \pi_n(n))$ is an exchangeable extension of $(\pi_n(1), \dots, \pi_n(k))$ meaning that the distribution of $(\pi_n(1), \dots, \pi_n(n))$ stays the same after any permutation. So by Theorem 1.1 of [25] we have that

$$\max_{1 \leq i_1, \dots, i_k \leq n} |\mathbb{P}(\pi_n(1) \leq i_1, \dots, \pi_n(k) \leq i_k) - \mathbb{E}[F_n(i_1) \cdots F_n(i_k)]| \leq C \frac{k}{n}$$

where C is a universal constant. Using Proposition III.B.3 and Corollary III.B.2 we find

$$\mathbb{E}[F_n(i_1) \cdots F_n(i_k)] = \frac{1}{n^k} \mathbb{E}_n[(S_{i_1} + i_1) \cdots (S_{i_k} + i_k)].$$


Indeed, Proposition III.B.3 implies that $(F_n(1), \dots, F_n(n))$ has the same distribution as $(G_n(1), \dots, G_n(n))$ with

$$G_n(i) := \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{r(j, T_{n+1}) \leq i}$$



where T_{n+1} is a uniform random tree of \mathfrak{C}_{n+1} . Then, Corollary III.B.2 shows that

$$G_n(i) = \frac{1}{n} \#\{1 \leq j \leq n : r(j, T_{n+1}) \leq i\} \stackrel{(d)}{=} \frac{1}{n} (S_i + i)$$

jointly for $i \in \{1, \dots, n\}$. This concludes the proof. 

E Sum and maximum of the first parking places

We begin this section with the proof of Corollary III.A.2. Then we finish by proving Proposition III.A.3.

Preuve du Corollaire III.A.2


(i) Recall that $(U_n(i))_{1 \leq i \leq n}$ are i.i.d. uniformly distributed in $\llbracket 1, n \rrbracket$. By the central limit theorem, the convergence

$$\sqrt{\frac{12}{k_n}} \left(\frac{U_n(1) + \dots + U_n(k_n)}{n} - \frac{k_n}{2} \right) \rightarrow \mathcal{N}(0, 1)$$

holds in distribution. Using the first item of Theorem III.A.1 we deduce that the total variation distance between the distributions of $\sum_{i=1}^{k_n} U_n(i)$ and $\sum_{i=1}^{k_n} \pi_n(i)$ tends to 0. Thus the above convergence still holds when $U_n(i)$ is replaced with $\pi_n(i)$.

(ii) Let $x > 0$, then

$$\mathbb{P} \left[k_n \left(1 - \frac{1}{n} \max\{U_n(1), \dots, U_n(k_n)\} \right) \geq x \right] = 0 \vee \frac{1}{n} \left[n \left(1 - \frac{x}{k_n} \right) \right]^{k_n} \xrightarrow{n \rightarrow \infty} e^{-x}.$$

Using Theorem III.A.1 (ii) we deduce that the above convergence still holds when $U_n(i)$ is replaced with $\pi_n(i)$. 

Preuve de la Proposition III.A.3

In this proof we write k instead of k_n to make notation lighter. For every $a \geq 0$:

$$\mathbb{P}(\pi_n(1), \dots, \pi_n(k) \leq n - a) = \frac{(n - k)!}{n!} \mathbb{E}_n(S_{n-a} + n - a)_k. \tag{E.1}$$

Indeed, following the same computation as in the proof of Proposition III.B.3, we have

$$\begin{aligned} \mathbb{P}(\pi_n(1), \dots, \pi_n(k) \leq n - a) &= \frac{(n - k)!}{n!} \sum_{\sigma \in \mathfrak{S}(k, n)} \mathbb{P}(\pi_n(\sigma(1)), \dots, \pi_n(\sigma(k)) \leq n - a) \\ &= \frac{(n - k)!}{n!} \mathbb{E} \left[\sum_{\sigma \in \mathfrak{S}(k, n)} \mathbb{1}_{r(\sigma(1), T_{n+1}), \dots, r(\sigma(k), T_{n+1}) \leq n - a} \right] \\ &= \frac{(n - k)!}{n!} \mathbb{E}[(X_1 + \dots + X_{n-a})_k] \end{aligned}$$

which leads to (E.1) since $X_1 + \dots + X_{n-a} = S_{n-a} + n - a$. Let τ_n be a Bienaymé-Galton-Watson tree with a critical Poisson offspring distribution μ conditionned on having n vertices and define S^n the associated Łukasiewicz path. More precisely, if v_1, \dots, v_n are the vertices of τ_n ordered according to the lexicogra-

phic order (see e.g. [78, Section 1.1]), then for all $0 \leq k \leq n$:

$$S_k^n := \#\{e : e \text{ is an edge adjacent to a vertex } v_i \text{ with } i \leq k\} - k.$$

In the previous definition, S_k^n is deduced from the first k vertices v_1, \dots, v_k but it is possible to see S_k^n in terms of the last $n - k$ vertices v_{k+1}, \dots, v_n :

$$S_k^n = n - 1 - k - \#\{e : e \text{ is an edge between two vertices } v_i \text{ and } v_j \text{ with } i, j > k\}.$$

It is known that S^{n+1} and S under \mathbb{P}_n have the same distribution. Thus, equality (E.1) can be rewritten in the following way


$$\mathbb{P}(\pi_n(1), \dots, \pi_n(k) \leq n - a) = \frac{(n - k)!}{n!} \mathbb{E}[(S_{n-a}^{n+1} + n - a)_k]. \quad (\text{E.2})$$

Let τ^* be the so-called Keřten's tree associated with μ (see e.g. [1, Section 2.3]). Denote by \leq the lexicographic order on the set of vertices of τ^* . It is always possible to find a unique infinite sequence u_1, u_2, \dots of distinct vertices of τ^* such that for all $i \geq 1$, $\{u : u \text{ is a vertex of } \tau^* \text{ such that } u_i \leq u\} = \{u_1, \dots, u_i\}$. In other words, u_1, u_2, \dots are the last vertices of τ^* for the lexicographic order, which, necessarily, lay on the right of the infinite spine. Similarly to the Łukasiewicz path we can define the quantity

$$S_a^* := a - \#\{e : e \text{ is an edge between two vertices } u_i \text{ and } u_j \text{ with } i, j \leq a + 1\}.$$

It is known that τ_n converges in distribution, for the local topology, towards τ^* (see e.g. [1, Section 3.3.5]). Making use of Skorokhod's representation theorem, suppose that the latter convergence holds almost surely. Thus, S_{n-a}^{n+1} converges almost surely towards S_a^* . Consequently, the convergence

$$\begin{aligned} \frac{(n - k)!}{n!} (S_{n-a}^{n+1} + n - a)_k &= \frac{(n - k)!}{(n - k + S_{n-a}^{n+1} - a)!} \frac{(n + S_{n-a}^{n+1} - a)!}{n!} \\ &\sim (n - cn)^{a - S_a^*} \frac{1}{n^{a - S_a^*}} \longrightarrow (1 - c)^{a - S_a^*} \end{aligned}$$

holds almost surely. Since the above sequence is bounded by 1 we deduce that the convergence of the expectation holds which concludes the proof. 



CHAPITRE

IV

PLUS GRAND ENSEMBLE
D'INDÉPENDANTS

Contenu du Chapitre

A Introduction	88
B Simply generated trees	90
C Main results	91
D Limit theorems for uniformly pointed simply generated trees	92
E Properties of the tricolouration	95
F Tricolouration of the infinite limiting trees	96
G Proof of Theorem IV.C.1	98



A Introduction

A subset S of vertices of a finite graph G is called an *independent set* if there is no pair of connected vertices in S . The *independence number* of G , denoted by $I(G)$, is the biggest cardinal of an independent set of G . The independence number is a well studied quantity in computational complexity theory. It is known that computing the independence number is NP-hard in general (see e.g. [56, Sec. 3.1.3]). A lot of work has been carried out to describe algorithms computing the independence number in general graphs [91, 102] and in special classes of graphs where the computational time can be decreased (e.g. cubic graphs [101], claw-free graphs [84], P_5 -free graphs [79]). The independence number has also received interest in combinatorics and in probability. Upper bounds have been found using probabilistic methods for cubic graphs [26]. Asymptotics have been found in certain classes of random trees (e.g. conditioned Bienaymé-Galton-Watson trees [44], simply generated trees [13], random recursive trees and binary search trees [54], and a wider class of random trees constructed from a Crump–Mode–Jagers branching process [70]). Finally we mention three articles giving applications of the independence number in scheduling theory [71], coding theory [29] and collusion detection in voting pools [12].

The goal of this article is to study the independence number of large simply generated trees, generalising some results of [40] and [13]. Simply generated trees are a wide class of random plane trees (i.e. rooted and ordered trees) introduced in [83] and encompass Bienaymé-Galton-Watson trees (BGW trees for short) conditioned to have a fixed number of vertices. Informally, a BGW tree with offspring distribution μ is a plane tree where vertices have an i.i.d. number of children with law μ . Various natural models of random trees are obtained with appropriate choices of the offspring distribution : e.g., uniform plane trees, uniform plane d -ary trees and uniform Cayley trees (see [68, Sec. 10]). In order to study the independence number, we will use a particular tricolouration of trees introduced in [105] and later studied in [39], [40] and [33]. This colouring is based on the notion of covering. A *covering* of a finite tree T is a subset of vertices S of T such that every edge of T is adjacent to a vertex of S . A *smallest covering* of T is a covering with minimal cardinality. In general, a tree has more than one smallest covering. For every vertex v of T we colour v in the following way :

- If v belongs to every smallest covering, we colour v in green.
- If v belongs to no smallest covering, we colour it in red.
- If v belongs to some smallest coverings but not all, we colour it in orange.

For a tree T , denote by $n_g(T)$, $n_o(T)$ and $n_r(T)$, respectively, the number of green, orange and red vertices in T . It has been noticed in [40] that the size of a smallest covering of a tree T is equal to $n_g(T) + n_o(T)/2$. Since the complementary of a smallest covering is an independent set of maximal size, the independence number of T is $n_r(T) + n_o(T)/2$. Actually, other statistics of the tree T can be expressed as a linear combination of $n_g(T)$, $n_o(T)$ and $n_r(T)$. For instance the *matching number* $M(T)$ (i.e. the maximum size of a partial vertex matching) is equal to the size of a smallest covering which is $n_g(T) + n_o(T)/2$. The *nullity* $N(T)$ (i.e. the kernel dimension of the adjacency matrix) is $n_r(T) - n_g(T)$. The *edge cover number* and the *clique cover number* also coincide with the independence number on trees (see [54]). Our main result (Theorem IV.C.1) concerns simply generated trees, but, to keep this introduction short, let us state here a particular case for critical BGW trees.

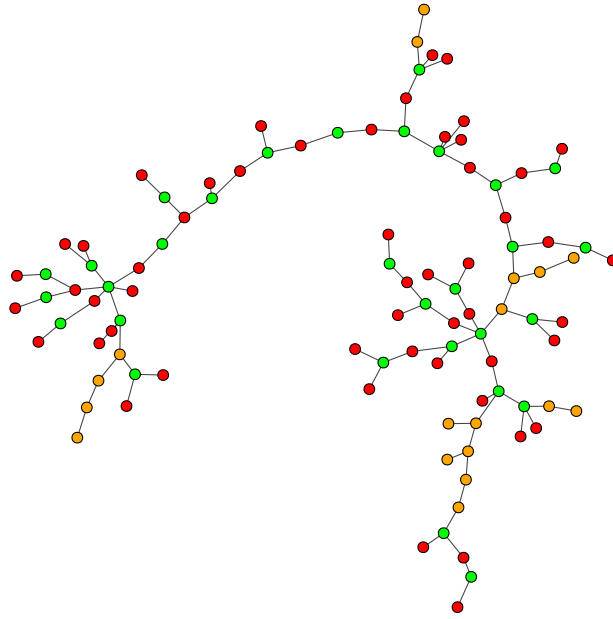


FIGURE A.1 – Tricolouration of a BGW tree with 100 vertices and a Poisson offspring distribution of parameter 1. The algorithm used to tricolour this tree can be found in [33, Appendix A].

Théorème IV.A.1

Let T_n be a Bienaymé-Galton-Watson tree with reproduction law μ , conditioned on having n vertices. Denote by $G(t) := \sum_{k=0}^{\infty} \mu_k t^k$ the generating function of μ and let q be the unique solution of $G(1 - q) = q$ in $[0, 1]$. Suppose that μ has mean 1, then, the following convergences hold in L^p for every $p > 0$:

$$\frac{n_g(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{1 - q + (1 - 2q)G'(1 - q)}{1 + G'(1 - q)}, \quad \frac{n_o(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{2qG'(1 - q)}{1 + G'(1 - q)},$$

$$\frac{n_r(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{q}{1 + G'(1 - q)}.$$

Explicit computations of the expected number of green, orange and red vertices have been carried out in the case of a uniform Cayley tree with fixed size in [40] using generating functions. This confirms the convergence, in mean value, of Theorem IV.A.1. Indeed, it is well known that a BGW tree with Poisson distribution of parameter 1 conditioned to have n vertices has the same law as a uniform Cayley tree with n vertices. To our knowledge, there is no estimates or asymptotics, other than [40], for the expected number of green, orange and red vertices in simply generated trees. As we said before, different quantities such as the independence number $I(T)$, the matching number $M(T)$ and the nullity $N(T)$ of a tree T can be expressed in terms of $n_g(T)$, $n_o(T)$ and $n_r(T)$, so Theorem IV.A.1 yields limit theorems for these in the case of critical BGW trees.

Corollaire IV.A.2

With the same notation and hypothesis as in Theorem IV.A.1, the following convergences hold in L^p for every $p > 0$:

$$\frac{I(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} q, \quad \frac{M(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 1 - q, \quad \frac{N(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 2q - 1.$$

With the same hypothesis, the authors of [44] show the convergence of $I(T_n)/n$ in probability towards q . Moreover, in [13] the convergence of the first and second moment of $I(T_n)/n$ is studied. More precisely, it is shown that $\mathbb{E}[I(T_n)] = nq + O(1)$ and $V(I(T_n)) = \nu n + O(1)$ for some constant ν . Therefore, the convergence of the independence number, in the settings of Corollary IV.A.2 is not new, but Corollary IV.C.2 generalises this convergence for simply generated trees that are not equivalent to conditioned critical BGW trees. The main tool to prove Theorem IV.C.1 is the use of limit theorems for uniformly pointed simply generated trees found in [97] (see Section D). In the first section we introduce simply generated trees. In the next section we state our main result in its most general form. To prove our main result, we first explain the limit theorems of [97] in the third section. The next two sections give properties of the tricolouration and describe how to colour the limiting trees. Finally, in the last section we prove Theorem IV.C.1.

B Simply generated trees

Let $\mathbf{w} := (w_i)_{i \geq 0}$ be a sequence of nonnegative weights. A *simply generated tree* having n vertices with weight sequence \mathbf{w} is a random plane tree T_n such that for every finite plane tree T ,

$$\mathbb{P}(T_n = T) = \frac{1}{Z_n} \left(\prod_{v \in T} w_{k_v} \right) \mathbb{1}_{|T|=n}$$

where k_v is the outdegree of the vertex v in T , $|T|$ is the number of vertices in T and Z_n is the normalising constant defined by

$$Z_n := \sum_{|T|=n} \prod_{v \in T} w_{k_v}.$$

Notice that, when the weight sequence \mathbf{w} is actually a probability sequence (i.e. the sum of the weights is equal to 1) then we recover the class of BGW trees. For T_n to be well defined, one needs Z_n to be nonzero. First of all, suppose that $w_0 > 0$ and $w_k > 0$ for some $k \geq 1$ otherwise $Z_n = 0$ for all $n \geq 1$. Let $\text{span}(\mathbf{w}) := \gcd\{i \geq 0 \mid w_i > 0\}$ (since $w_k > 0$ this quantity is well defined). The following result, found for instance in [68, Cor. 15.6], characterises the n 's such that $Z_n > 0$:

Lemme IV.B.1 : Janson [68]

If $Z_n > 0$ then $n \equiv 1 \pmod{\text{span}(\mathbf{w})}$. Conversely, there exists n_0 such that for all $n \geq n_0$ satisfying $n \equiv 1 \pmod{\text{span}(\mathbf{w})}$, $Z_n > 0$.

Throughout this chapter and in Theorem IV.C.1, we suppose that $w_0 > 0$, $w_k > 0$ for some $k \geq 1$ and that all the n 's appearing satisfy $n \geq n_0$ and $n \equiv 1 \pmod{\text{span}(\mathbf{w})}$.

Let $\rho \in [0, +\infty]$ be the radius of convergence of the generating series

$$\phi(x) := \sum_{i \geq 0} w_i x^i.$$

It is shown in [68, Lemma 3.1] that, if $\rho > 0$, then the function defined by

$$\psi(x) := \frac{x\phi'(x)}{\phi(x)}$$

is increasing on $[0, \rho)$ and we can define $\nu := \lim_{x \rightarrow \rho} \psi(x) \in (0, +\infty]$. We distinguish three different regimes :

- Regime 1 when $\rho > 0$ and $\nu \geq 1$. In this case there is a unique $\tau \in [0, \rho]$ such that $\tau < +\infty$ and $\psi(\tau) = 1$.
- Regime 2 when $\rho > 0$ and $0 < \nu < 1$. In this case $\rho < +\infty$ and we set $\tau := \rho$.
- Regime 3 when $\rho = 0$.

In regime 1 and 2 we can define a probability function given by

$$\pi_k := \frac{\tau^k w_k}{\phi(\tau)}. \quad (\text{B.1})$$

The associated mean and generating function are respectively given by

$$m := \min(1, \nu) \quad \text{and} \quad G(x) := \frac{\phi(\tau x)}{\phi(x)}. \quad (\text{B.2})$$

An important result of [68] is that, T_n , in regime 1 or 2, has the same law as a BGW tree with reproduction law π conditioned to have n vertices. In regime 3, T_n is not distributed like a conditioned BGW tree. Note that a critical or super-critical BGW tree or a BGW tree with a reproduction law with infinite mean, conditioned to have n vertices, always lays in regime 1. Moreover for a critical BGW tree with reproduction law μ , the probability π is the same as μ . A sub-critical BGW tree, conditioned to have n vertices, is either in regime 1 or in regime 2. We define *complete condensation* to be the condition :

$$\Delta(T_n) = (1 - m)n + n\mathcal{E}_n \quad (\text{B.3})$$

where $\Delta(T_n)$ is the maximum degree of a vertex of T_n and \mathcal{E}_n is a random variable converging in probability towards 0. For instance, complete condensation happens in regime 2 when there exists $\theta > 1$ and a slowly varying function ℓ such that $\pi_k = \ell(k)k^{-(1+\theta)}$ (see [76]). Complete condensation also happens in regime 3 for the weight sequence $w_k = k!^\alpha$ for $\alpha > 0$ (see [68, Ex. 19.36]).

C Main results

In this section we state our main results. We keep all the notations and assumptions of Section B.

Théorème IV.C.1

Let T_n be a simply generated tree with n vertices according to the weight sequence $\mathbf{w} = (w_i)_{i \geq 0}$. Recall that, in regime 1 and 2, G denotes the generating function of π defined in (B.1). Let q be the unique solution of $q = G(1 - q)$ in $[0, 1]$.

1. In regime 1 and in regime 2 with complete condensation (meaning that (B.3) is satisfied), the following convergences hold in L^p for every $p > 0$:

$$\frac{n_g(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{1 - q + (1 - 2q)G'(1 - q)}{1 + G'(1 - q)}, \quad \frac{n_o(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{2qG'(1 - q)}{1 + G'(1 - q)},$$

$$\frac{n_r(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} \frac{q}{1 + G'(1 - q)}.$$

2. In regime 3 with complete condensation, the following convergences hold in L^p for every $p > 0$:

$$\frac{n_g(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 0, \quad \frac{n_o(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 0, \quad \frac{n_r(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 1.$$

Corollaire IV.C.2

We keep the same notation and hypothesis as in Theorem IV.C.1. Recall that $I(T_n)$, $M(T_n)$ and $N(T_n)$ are, respectively, the independence number, the matching number and the nullity of T_n .

1. In regime 1 and in regime 2 with complete condensation, the following convergences hold in L^p for every $p > 0$:

$$\frac{I(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} q, \quad \frac{M(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 1 - q, \quad \frac{N(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 2q - 1.$$

2. In regime 3 with complete condensation, the following convergences hold in L^p for every $p > 0$:

$$\frac{I(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 1, \quad \frac{M(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 0, \quad \frac{N(T_n)}{n} \xrightarrow[n \rightarrow \infty]{L^p} 1.$$

D Limit theorems for uniformly pointed simply generated trees

In this section we explain the results proved in [97] which will be our basic tool to prove Theorem IV.C.1. All the proofs and details of this section can be found in the above mentioned article. As said in the introduction, these results are limit theorems for uniformly pointed simply generated trees. A *pointed tree* is simply a couple (T, v) with a plane tree (i.e. rooted and ordered tree) T and a distinguished vertex v of T . A uniformly pointed simply generated tree is a couple (T_n, v_n) where T_n is a simply generated tree with n vertices and v_n is a distinguished vertex chosen uniformly at random among the n vertices of T_n . Basically, in regime 1 and in regime 2 and 3 with complete condensation, the local tree structure around v_n converges towards an infinite random tree which depends only on the regime. To formally define the notion of convergence used here, one needs to consider T_n to be a subtree of a big ambient tree denoted by $\mathcal{U}_\infty^\bullet$. Every plane tree (e.g. T_n) is considered, by definition, to be rooted, however we will encounter some infinite trees without any root which is



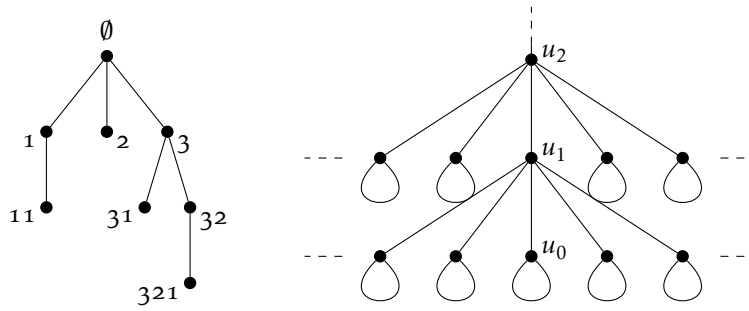


FIGURE D.1 – On the left, a subtree of the Ulam-Harris tree \mathcal{U}_∞ . On the right, the tree $\mathcal{U}_\infty^\bullet$. Each loop represents a copy of the Ulam-Harris tree.

unusual in the classic framework of plane trees. Let

$$\mathcal{V}_\infty := \{\emptyset\} \cup \bigcup_{n \geq 1} (\mathbb{N}^*)^n$$

be the set of words (empty word included) formed in the alphabet $\mathbb{N}^* = \{1, 2, 3, \dots\}$. Usually, plane trees are defined as subtrees of the so-called *Ulam-Harris tree*, denoted here by \mathcal{U}_∞ , which is the tree with vertex set \mathcal{V}_∞ and edge set $\{(a_1 \dots a_{n-1}; a_1 \dots a_n) \mid \forall n, a_1, \dots, a_n \in \mathbb{N}^*\}$. With this definition, all the plane trees have a root which is a common ancestor to every vertex of the tree (it is the vertex designated by the empty word \emptyset). However, in regime 1, the root of T_n is, in a local point of view, at infinite distance from the distinguished vertex v_n . It suggests that the local limit of T_n around v_n has an infinite spine of ancestors and therefore, has no root. This is why we need a more general framework than the usual one for plane trees. Here we explain informally the construction of $\mathcal{U}_\infty^\bullet$. Let u_0, u_1, \dots be vertices, in the plane, lined up to form an infinite connected spine. Each vertex u_i with $i > 0$ gets an infinite countable number of children on the left and on the right of its child u_{i-1} . Then, all the leaves of the current tree (u_0 included) give birth to the Ulam-Harris tree \mathcal{U}_∞ . The tree we obtain from this construction is denoted by $\mathcal{U}_\infty^\bullet$ and its set of vertices is denoted by $\mathcal{V}_\infty^\bullet$ (see Figure D.1). To formally define this tree one could start by creating the vertex set $\mathcal{V}_\infty^\bullet$ as a subset of $\mathbb{N} \times \mathbb{Z} \times \mathcal{V}_\infty$ and then describing the edge set. However we think that the above informal construction is enough for our purpose. As we said, we want to represent T_n as a subtree of $\mathcal{U}_\infty^\bullet$. First we make clear what we call a subtree of \mathcal{U}_∞ and $\mathcal{U}_\infty^\bullet$. Denote by \mathcal{E}_∞ the edge set of \mathcal{U}_∞ and $\mathcal{E}_\infty^\bullet$ the edge set of $\mathcal{U}_\infty^\bullet$.

Définition IV.D.1

A *subtree* t of \mathcal{U}_∞ is a tree with vertex set included in \mathcal{V}_∞ and edge set included in \mathcal{E}_∞ , such that the vertex \emptyset belongs to t and such that there are no holes in t , meaning that : if $v = (a_1 \dots a_n)$ is a vertex of t with $a_n > 0$ then $(a_1 \dots a_{n-1})$ is also a vertex of t . Similarly a *subtree* t of $\mathcal{U}_\infty^\bullet$ is a tree with vertex set included in $\mathcal{V}_\infty^\bullet$ and edge set included in $\mathcal{E}_\infty^\bullet$, such that the vertex u_0 belongs to t and such that there are no holes in t (see Figure G.1). A subtree of $\mathcal{U}_\infty^\bullet$ is *rooted* if the set $\{k \geq 0 \mid u_k \in t\}$ is finite. In this case the vertex u_k with maximal k in t is called the *root* of t .

The representation of T_n as a subtree of $\mathcal{U}_\infty^\bullet$ will obviously depend on the distinguished vertex v_n since we want to look at the local structure around this vertex. More precisely, let T be a plane tree and v be a vertex of T . We identify the distinguished vertex v with the element u_0 and the root of T is identified with the element

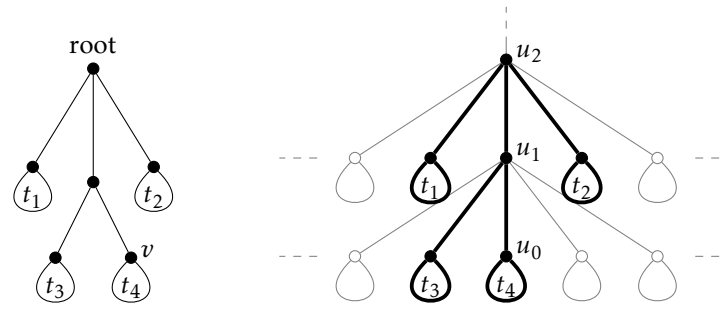


FIGURE D.2 – On the left, a finite tree T with a distinguished vertex v at distance 2 from the root. The loops t_1, t_2, t_3 and t_4 represent subtrees of T which can be seen as subtrees of the Ulam-Harris tree \mathcal{U}_∞ . On the right, the representation of T as a subtree of $\mathcal{U}_\infty^\bullet$.

u_h where h is the graph distance between the root and v in T (it is the height of v). All the other vertices of T are identified such that the plane order is preserved (See Figure D.2). We denote by (T, v) the subtree of $\mathcal{U}_\infty^\bullet$ obtained this way. We can now formally define the notion of convergence we use.

Définition IV.D.2

Let t_n and t be subtrees of $\mathcal{U}_\infty^\bullet$ for all n . We say that (t_n) converges towards t and write $t_n \rightarrow t$ if for all $v \in \mathcal{V}_\infty^\bullet$,

$$\mathbb{1}_{v \in t_n} \xrightarrow{n \rightarrow \infty} \mathbb{1}_{v \in t}.$$

This notion of convergence induces a topology that is metrizable and compact over the set of subtrees of $\mathcal{U}_\infty^\bullet$. Before stating the limit theorems, one needs to define the limiting trees T_1^* , T_2^* and T_3^* (seen as subtrees of $\mathcal{U}_\infty^\bullet$) that correspond, respectively, to regime 1, 2 and 3. Let T be a BGW tree with reproduction law π , given by (B.1), in regime 1 or 2. Let $\hat{\pi}$ be the probability measure on $\mathbb{N} \cup \{\infty\}$ given by

$$\hat{\pi}_k = k\pi_k \quad \forall k \in \mathbb{N} \quad \text{and} \quad \hat{\pi}_\infty = 1 - m$$

where $m := \min(1, \nu)$ is the mean of π .

- First we define T_1^* in regime 1. Notice that in this case $\hat{\pi}_\infty = 0$. We attach to u_0 an independent copy of T . For $k \geq 1$, u_k receives offspring according to an independent copy of $\hat{\pi}$. Then u_{k-1} is identified with a child of u_k chosen uniformly at random. Finally, we attach an independent copy of T to all the children of u_k , except u_{k-1} (see Figure D.3).
- Now we define T_2^* in regime 2. In this case $\hat{\pi}_\infty > 0$. We attach to u_0 an independent copy of T . For $k \geq 1$, u_k receives offspring according to an independent copy of $\hat{\pi}$. Notice that almost surely, there exist $1 \leq i < j$ two integers such that $u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_{j-1}$ have a finite number of children and u_i and u_j have an infinite number of children. For every $k \in \{1, \dots, i-1, i+1, \dots, j-1\}$, u_{k-1} is identified with a child of u_k chosen uniformly at random, while u_i gets infinitely many children on the left and the right of its child u_{i-1} . Finally, for all $k \geq 1$, we attach an independent copy of T to all the children of u_k , except u_{k-1} . The tree T_2^* is the tree obtained by keeping all the descendants of u_{j-1} (see Figure D.3).
- Finally, T_3^* is simply composed of the vertex u_1 having infinitely many children on the left and on the right

of u_0 , all of them, including u_0 , being leaves (see Figure D.3).

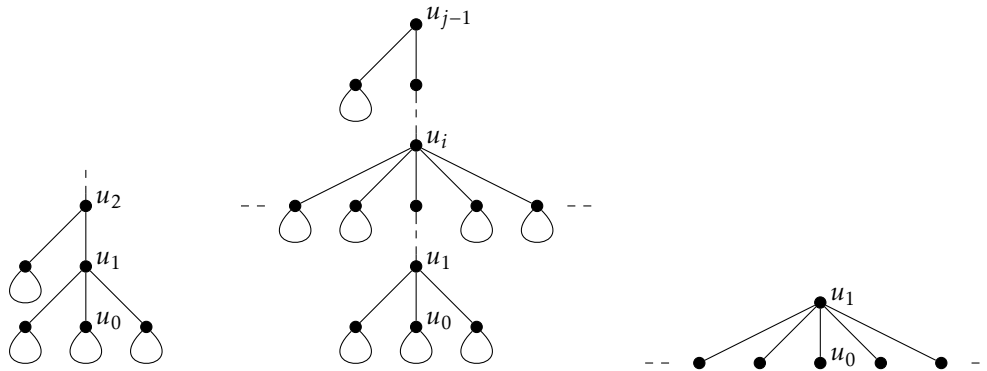


FIGURE D.3 – Representation of the trees T_1^* , T_2^* and T_3^* , respectively, from left to right. Each loop represents a copy of a Bienaymé–Galton–Watson tree of reproduction law π .

Théorème IV.D.3 : Stufler [97]

Let (T_n, v_n) be a uniformly pointed simply generated tree with n vertices. Suppose that we are in regime $i = 1$ or in regime $i \in \{2, 3\}$ with complete condensation (meaning that (B.3) is satisfied). Then the convergence

$$(T_n, v_n) \xrightarrow[n \rightarrow \infty]{(d)} T_i^*$$

holds in distribution for the topology induces by the convergence of Definition IV.D.2.

E Properties of the tricolouration

In this section, we look at some general properties of the tricolouration defined in the introduction that will be essential to prove our main result. Let T_1, \dots, T_n be n rooted finite trees. We define $T_1 * \dots * T_n$ the rooted tree obtained by creating an edge between each root of T_1, \dots, T_n and a new vertex which will be the root of $T_1 * \dots * T_n$. In particular the number of vertices $\#V(T_1 * \dots * T_n)$ equals $1 + \#V(T_1) + \dots + \#V(T_n)$. And the number of edges $\#E(T_1 * \dots * T_n)$ equals $n + \#E(T_1) + \dots + \#E(T_n)$. We say that a rooted tree has colour c if the root has colour c .

Lemme IV.E.1

Let T_1, \dots, T_n be n rooted finite trees. Set $T := T_1 * \dots * T_n$, then

1. T is red if T_1, \dots, T_n are all non-red.
2. T is orange if exactly one tree among T_1, \dots, T_n is red.
3. T is green if two or more trees among T_1, \dots, T_n are red.

Preuve

Denote by $C(T)$ the size of a smallest covering of T . Notice that $C(T)$ is equal to $C(T_1) + \dots + C(T_n) + \Delta$ with $\Delta \in \{0, 1\}$. More precisely, $C(T) = C(T_1) + \dots + C(T_n)$ if and only if all the T_i 's are non-red.

1. Suppose that T_1, \dots, T_n are all non-red. We can take a smallest covering for each T_i such that the root of T_i is included in the covering. Then the union of these coverings gives a smallest covering of T . Moreover we can see that all the smallest coverings of T are obtained this way. Thus T is red.
2. Suppose that T_1 is red and T_2, \dots, T_n are all non-red. We can take a smallest covering for each T_i in addition to the root of T . This gives a smallest covering of T , so T is either green or orange. We can also take a smallest covering for each $T_i, i > 1$, such that the root of T_i is included in the covering, a smallest covering of T_1 and the root of T_1 . This also gives a smallest covering of T . Thus T is orange.
3. Suppose that T_1 and T_2 are red. As for the previous case, we can take a smallest covering for each T_i in addition to the root of T . This gives a smallest covering for T . But, as opposed to the previous case, all the smallest coverings of T are obtained this way. Thus T is green.



If T_1, T_2 are finite trees and v_1, v_2 are vertices of, respectively, T_1 and T_2 , then we denote by $(T_1, v_1) * (T_2, v_2)$ the tree obtained from T_1 and T_2 by drawing an edge between v_1 and v_2 .

Lemme IV.E.2

With the same notation as above, set $T := (T_1, v_1) * (T_2, v_2)$. If v_1 is green in the tricolouration of T_1 then, the colour of every vertex v in the tricolouration of T is just the same as its colour in the tricolouration of T_1 (if v is a vertex of T_1) or T_2 (if v is a vertex of T_2). In other words, the tricoloured tree T is simply obtained by drawing an edge between v_1 and v_2 and keeping the colours of T_1 and T_2 .

Preuve

Notice that, since v_1 is green in the tricolouration of T_1 , a smallest covering of T_1 combined with a smallest covering of T_2 gives a smallest covering of T . Conversely a smallest covering of T is necessarily obtained by combining a smallest covering of T_1 and T_2 .



F Tricolouration of the infinite limiting trees

In this section, we extend the definition of the tricolouration given in the introduction to the random infinite limiting trees T_1^*, T_2^* and T_3^* defined in Section D. The initial definition applies only to finite trees since a covering of smallest size only makes sense in this context. Even though it seems not obvious to find a satisfactory definition of "smallest covering" for an infinite tree, it is still possible to describe a canonical way to tricolour the trees T_1^*, T_2^* and T_3^* using the properties found in Section E.

Let t be a subtree of $\mathcal{U}_\infty^\bullet$ (finite or not) such that for every $k \geq 0$ and for every child v of u_k , distinct from u_{k-1} , v has a finite number of descendants. In other words, for all $k \geq 0$, all the children of u_k , distinct from

u_{k-1} , are roots of finite trees. Since those trees are finite, it makes sense to consider their tricolouration. A *good* vertex of t is a vertex u_k with $k \geq 0$ such that, at least two of its children, distinct from u_{k-1} , are red in the tricolouration of the finite subtree they produce. If t is finite, then, from Lemma IV.E.1, a good vertex is a green vertex for the tricolouration of t . Notice that T_1^* , T_2^* and T_3^* satisfy the same hypothesis as t almost surely.

— We begin with the definition of the tricolouration of T_1^* . Almost surely, there exists an increasing sequence $(k_i)_i$ such that for all i , u_{k_i} is good. For all i , all the vertices below u_{k_i} (u_{k_i} included) get the same colour in T_1^* as their colour in the tricolouration of the subtree rooted at u_{k_i} (which is a finite tree). Notice that, from Lemma IV.E.1, u_{k_i} gets necessarily the colour green. Lemma IV.E.2 ensures that this way of colouring is consistent when taking larger i .

— For T_2^* we colour the unique vertex with infinite degree in green. Then, by cutting this vertex from T_2^* we obtain a (infinite) forest of finite trees who gets their induced tricolouration. Notice that, almost surely, the vertex with infinite degree is good.

— Lastly, all the leaves of T_3^* are coloured in red and the root u_1 is coloured in green.

We finish this section with the following lemma which explicitly gives the colour distribution of the vertex u_0 in T_i^* . This lemma will be useful when proving Theorem IV.C.1.

Lemme IV.E.1

Let $p_i(c)$ be the probability that u_0 has colour c in T_i^* .

1. In regime $i = 1$ and $i = 2$ with complete condensation we have

$$p_i(\text{green}) = \frac{1 - q + (1 - 2q)G'(1 - q)}{1 + G'(1 - q)}, \quad p_i(\text{orange}) = \frac{2qG'(1 - q)}{1 + G'(1 - q)},$$

$$p_i(\text{red}) = \frac{q}{1 + G'(1 - q)}.$$

2. In regime 3 with complete condensation we have that $p_3(\text{red}) = 1$.

Preuve

The case of regime 3 is obvious, let us focus on regime 1 and 2. Let T be a BGW tree with reproduction law π . Denote by q the probability that the root of T is red. From Lemma IV.E.1 we deduce that

$$q = \sum_{k \geq 0} \pi_k (1 - q)^k = G(1 - q).$$

Let \tilde{T} be the tree obtained from T_i^* by cutting the edge between u_0 and u_1 and keeping the component containing u_1 . Let \tilde{q} be the probability that u_1 is red in \tilde{T} . Then, from Lemma IV.E.1 again,

$$\tilde{q} = \sum_{k \geq 1} k \pi_k (1 - q)^{k-1} (1 - \tilde{q}) = (1 - \tilde{q})G'(1 - q).$$

Finally,

$$p_i(\text{red}) = \sum_{k \geq 0} \pi_k (1 - q)^k (1 - \tilde{q}) = \frac{q}{1 + G'(1 - q)}.$$

And

$$p_i(\text{orange}) = \sum_{k \geq 0} \pi_k (1 - q)^k \tilde{q} + \sum_{k \geq 1} k \pi_k (1 - q)^{k-1} q (1 - \tilde{q}) = \frac{2qG'(1 - q)}{1 + G'(1 - q)}.$$

Finally we deduce the value $p_i(\text{green})$ by the law of total probability.



G Proof of Theorem IV.C.1

All this section is devoted to the proof of Theorem IV.C.1. We keep all the notation of Theorem IV.C.1 and suppose that we are in regime $i = 1$ or in regime $i \in \{2, 3\}$ with complete condensation. Let c be a colour in {red, green, orange}. Recall that $p_i(c)$ is the probability that the vertex u_0 has colour c in the tree T_i^* in regime i . The idea is to prove the convergence of the first two moments of $n_c(T_n)/n$, namely

$$\frac{1}{n} \mathbb{E}[n_c(T_n)] \xrightarrow{n \rightarrow \infty} p_i(c) \quad \text{and} \quad \frac{1}{n^2} \mathbb{E}[n_c(T_n)^2] \xrightarrow{n \rightarrow \infty} p_i(c)^2.$$

Then, using Lemma IV.G.1, we will conclude that $n_c(T_n)/n$ converges in L^p towards $p_i(c)$ for all $p > 0$. Actually, the convergence of the second moment won't be required in regime 3. Recall that the explicit computation of $p_i(c)$ can be found in Lemma IV.F.1.

Lemme IV.G.1

Let (X_n) be a sequence of random variables with values in $[0, 1]$, and $\alpha \in [0, 1]$. Suppose that one of the following condition is satisfied.

- The convergences $\mathbb{E}[X_n] \rightarrow \alpha$ and $\mathbb{E}[X_n^2] \rightarrow \alpha^2$ hold when $n \rightarrow \infty$.
- The convergence $\mathbb{E}[X_n] \rightarrow \alpha$ holds when $n \rightarrow \infty$ and $\alpha = 1$.

Then for all $p > 0$, (X_n) converges towards α in L^p .

Preuve du Lemme IV.G.1

First, we show that (X_n) converges towards α in probability. Let $\varepsilon > 0$. In the first case we use Markov's inequality which gives

$$\mathbb{P}(|X_n - \alpha| \geq \varepsilon) \leq \frac{\mathbb{E}[(X_n - \alpha)^2]}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

In the second case we notice that

$$\mathbb{E}[X_n] \leq (1 - \varepsilon)\mathbb{P}(X_n \leq 1 - \varepsilon) + \mathbb{P}(X_n > 1 - \varepsilon) = 1 - \varepsilon\mathbb{P}(X_n \leq 1 - \varepsilon).$$

Consequently $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq 1 - \varepsilon) = 0$ and the convergence in probability is shown in both cases. Second, let ℓ be an accumulation point of the sequence $(\mathbb{E}[|X_n - \alpha|^p])_n$ and $(n_k)_k$ be an extraction such that the convergence to ℓ occurs. From $(X_{n_k})_k$ we can extract a subsequence that converges almost surely to α . From the dominated convergence theorem we deduce that $\ell = 0$.



Let v_n, v'_n be vertices chosen independently and uniformly in T_n . Notice that

$$\begin{aligned} \frac{1}{n} \mathbb{E}[n_c(T_n)] &= \mathbb{P}(v_n \text{ has colour } c \text{ in } T_n) \quad \text{and} \\ \frac{1}{n^2} \mathbb{E}[n_c(T_n)^2] &= \mathbb{P}(v_n \text{ and } v'_n \text{ have colour } c \text{ in } T_n). \end{aligned}$$



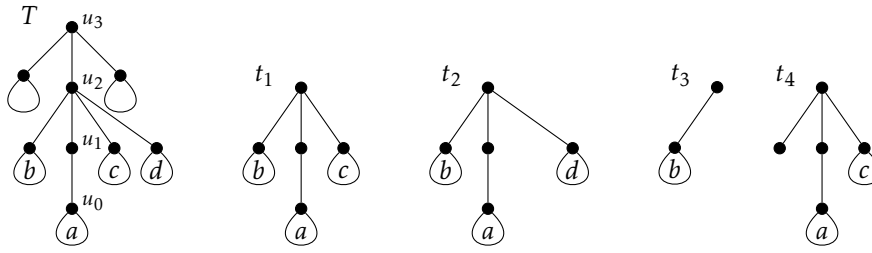


FIGURE G.1 – Illustration of Definitions IV.D.1 and IV.G.2. Only the tree t_1 satisfies $t_1 \leq T$. The tree t_3 is not even a subtree of $\mathcal{U}_\infty^\bullet$ since it doesn't contain u_0 . The tree t_2 is not a subtree of $\mathcal{U}_\infty^\bullet$ either since it has a hole between u_1 and the rightmost child of u_2 . Finally t_4 is a subtree of $\mathcal{U}_\infty^\bullet$ but doesn't satisfy $t_4 \leq T$ because the descendants of the left child of u_2 are missing.

Convergence of the first moment First we prove the convergence of the first moment. Recall the notation of Section F when defining the tricolouration of the infinite trees T_1^* , T_2^* and T_3^* . Denote by $k \geq 0$ the first positive integer such that u_k is good ($k = 1$ almost surely in regime 3). Let τ_i^* be the subtree of $\mathcal{U}_\infty^\bullet$ obtained from T_i^* by cutting the edge between u_k and u_{k+1} and keeping the component containing u_0 ($\tau_3^* = T_3^*$ in regime 3). Note that, by construction, the tricolouration of τ_i^* is the restriction of its tricolouration in T_i^* and that u_k is green. The following definition introduces a useful order relation between trees.

Définition IV.G.2

Let T and t be subtrees of $\mathcal{U}_\infty^\bullet$ such that t is rooted at u_j for some $j \geq 0$. Suppose that u_j is also a vertex of T and denote by E_j the set of edges of T adjacent to u_j . We write $t \leq T$ if there exists a subset $e_j \subset E_j$ such that t is the tree obtained from T by cutting all the edges from e_j and keeping the component containing u_j (see Figure G.1).

Let \mathcal{F} be the set of rooted subtrees t of $\mathcal{U}_\infty^\bullet$ such that the root $u_j \in t$ is the only good vertex of t . For $t \in \mathcal{F}$ such that the root u_j of t has finite degree, denote by $v_\ell(t)$ (resp. $v_r(t)$) the leftmost (resp. rightmost) child of the root u_j of t . Let \mathcal{F}_0 be the set of elements $t \in \mathcal{F}$ such that : the root u_j of t has finite degree; u_j has exactly two red neighbors distinct from u_{j-1} ; and ($v_\ell(t) = u_{j-1}$ or $v_\ell(t)$ is red) and ($v_r(t) = u_{j-1}$ or $v_r(t)$ is red). Notice that if t_1 and t_2 are distinct elements of \mathcal{F}_0 , then we can't have $t_1 \leq t_2$ nor $t_2 \leq t_1$. Moreover for every $t_1 \in \mathcal{F}$ there exists a unique $t_2 \in \mathcal{F}_0$ such that $t_2 \leq t_1$. In other words \mathcal{F}_0 is the set of equivalence classes for the equivalence relation $t_1 \sim t_2$ iff $t_1 \leq t_2$ or $t_2 \leq t_1$. Notice that almost surely $\tau_i^* \in \mathcal{F}$.

Fix $\varepsilon > 0$. Let \mathcal{T} be a finite subset of \mathcal{F}_0 such that the event $\{\mathcal{T} \leq \tau_i^*\} := \{\exists t \in \mathcal{T}, t \leq \tau_i^*\}$ happens with probability at least $1 - \varepsilon$. Let $\mathcal{T}(c)$ be the set of trees $t \in \mathcal{T}$ such that u_0 has colour c in t . Remember that we see (T_n, v_n) as a subtree of $\mathcal{U}_\infty^\bullet$. For all $t \in \mathcal{T}$, define the event $A_n(t) := \{t \leq (T_n, v_n)\}$. Using Theorem IV.D.3, we have that for all $t \in \mathcal{T}$

$$\mathbb{P}(A_n(t)) \xrightarrow{n \rightarrow \infty} \mathbb{P}(t \leq T_i^*). \tag{G.1}$$

The properties of τ_i^* and $t \in \mathcal{T}$ imply that $t \leq T_i^*$ if and only if $t \leq \tau_i^*$. Thus

$$\mathbb{P}(t \leq T_i^*) = \mathbb{P}(t \leq \tau_i^*).$$

Denote by E_n the event $\cup_{t \in \mathcal{T}} A_n(t)$. Notice that the event $E_n \cap \{v_n \text{ has colour } c \text{ in } T_n\}$ is equal to the event

$\cup_{t \in \mathcal{T}(c)} A_n(t)$. It is a consequence of Lemma IV.E.1 and IV.E.2. Notice also that for t_1, t_2 distinct trees of \mathcal{T} , $A_n(t_1)$ and $A_n(t_2)$ are disjoint for all n . Consequently,

$$\mathbb{P}(\{v_n \text{ has colour } c \text{ in } T_n\} \cap E_n) = \sum_{t \in \mathcal{T}(c)} \mathbb{P}(A_n(t)) \xrightarrow{n \rightarrow \infty} \mathbb{P}(\mathcal{T}(c) \leq \tau_i^*).$$

Since u_0 has colour c in τ_i^* if and only if u_0 has colour c in T_i^* ,

$$\mathbb{P}(\mathcal{T}(c) \leq \tau_i^*) = \mathbb{P}(u_0 \text{ has colour } c \text{ in } T_i^* \text{ and } \mathcal{T} \leq \tau_i^*) \in [p_1(c) - \varepsilon, p_1(c) + \varepsilon].$$

Finally, using Theorem IV.D.3 again, one has

$$\mathbb{P}(E_n) \xrightarrow{n \rightarrow \infty} \mathbb{P}(\mathcal{T} \leq \tau_i^*) \geq 1 - \varepsilon.$$

The convergence $\mathbb{E}[n_c(T_n)/n] \rightarrow p_1(c)$ readily follows.

Convergence of the second moment in regime 1 and 2 The next step is to show convergence for the second moment in regime $i = 1$ or $i = 2$ with complete condensation. Let c' be another colour in {red, green, orange}.

We will actually show that

$$\frac{1}{n^2} \mathbb{E}[n_c(T_n)n_{c'}(T_n)] \xrightarrow{n \rightarrow \infty} p_i(c)p_i(c').$$

We keep the notation of the previous part which shows the convergence of the first moment. For all $t \in \mathcal{T}$, let $A'_n(t) := \{t \leq (T_n, v'_n)\}$ and $E'_n := \cup_{t \in \mathcal{T}} A'_n(t)$. We have,

$$\mathbb{P}(\{v_n \text{ has colour } c \text{ in } T_n\} \cap E_n \cap E'_n) = \sum_{t \in \mathcal{T}(c)} \sum_{t' \in \mathcal{T}(c')} \mathbb{P}(A_n(t) \cap A'_n(t')). \tag{G.2}$$

Fix $t, t' \in \mathcal{T}$. Recall that the trees t, t' and T_n are rooted plane trees, thus we can consider their so-called Łukasiewicz walk. More precisely, let T be a rooted plane tree with n vertices and w be a vertex of T . Denote by $\ell(w, T)$ the rank of w in T for the lexicographic order. Equivalently, w is the $\ell(w, T)$ -th vertex of T explored by the depth first search starting from the root of T . Let w_1, \dots, w_n be the vertices of T ordered according to the lexicographic order (so $\ell(w_i, T) = i$ for all i). The Łukasiewicz walk associated with T is the sequence $(s_k)_{1 \leq k \leq n}$ such that $s_0 = 0$ and $s_k - s_{k-1} + 1$ is the out-degree of w_k for all $k \in \{1, \dots, n\}$. An important property of the Łukasiewicz walk is that it uniquely encodes its tree, meaning that the tree T can be retrieved from $(s_k)_{1 \leq k \leq n}$. Let $(S_k^{(n)})_{0 \leq k \leq n}$ be the Łukasiewicz walk associated with the tree T_n . Let X_1, \dots, X_n, \dots be i.i.d random variables such that $\mathbb{P}(X_1 = m) = \pi_{m+1}$ for all integer $m \geq -1$ and set $S_k := \sum_{i=1}^k X_i$ for all $k \geq 0$. It is well known that, the random walk $(S_k)_{0 \leq k \leq n}$, starting at 0 and conditioned on reaching -1 for the first time at time n , has the same law as $(S_k^{(n)})_{0 \leq k \leq n}$. Let $m := |t|$ be the number of vertices of t and $m' := |t'|$. Let $(s_k)_{0 \leq k \leq m}$ and $(s'_k)_{0 \leq k \leq m'}$ be, respectively, the Łukasiewicz walks associated with t and t' and denote by $x_k := s_{k+1} - s_k$ and $x'_k := s'_{k+1} - s'_k$ the associated steps. Write $k_0 := \ell(u_0, t)$, $k'_0 := \ell(u_0, t')$, $i_n := \ell(v_n, T_n)$ and $i'_n := \ell(v'_n, T_n)$. The indices i_n and i'_n are independent random elements of $\{1, \dots, n\}$ with uniform distribution. The event $A_n(t)$ happens if and only if the Łukasiewicz walk $(S_k^{(n)})_{0 \leq k \leq n}$ coincides with $(s_k)_{0 \leq k \leq m}$, up to a vertical shifting, on the interval $\llbracket i_n - k_0, i_n + m - k_0 \rrbracket$. The same goes for $A'_n(t')$. More precisely

$$A_n(t) \cap A'_n(t') = \{X_{i+i_n-k_0}^{(n)} = x_i \forall i \in \llbracket 1, m \rrbracket \text{ and } X_{i+i'_n-k'_0}^{(n)} = x'_i \forall i \in \llbracket 1, m' \rrbracket\}.$$

Applying the reverse Vervaat transform, we can change the initial excursion type conditioning into a bridge type conditioning (see e.g. [88, Sec. 6.1]). Namely

$$\mathbb{P}(A_n(t) \cap A'_n(t')) = \mathbb{P}(X_{i+i_n-k_0} = x_i \forall i \in \llbracket 1, m \rrbracket \text{ and } X_{i+i'_n-k'_0} = x'_i \forall i \in \llbracket 1, m' \rrbracket \mid X_n = -1).$$

Denote by D_n the event $\{[i_n - k_0 + 1, i_n - k_0 + m] \cap [i'_n - k'_0 + 1, i'_n - k'_0 + m'] = \emptyset\}$. This event has a probability tending to 1 and one can see that

$$\mathbb{P}(A_n(t) \cap A'_n(t') \cap D_n) = \mathbb{P}(X_i = x_i \forall i \in [1, m] \text{ and } X_{i+m} = x'_i \forall i \in [1, m'] \mid X_n = -1) \mathbb{P}(D_n).$$

According to [68, Thm. 11.7] the steps $X_1, \dots, X_{m+m'}$ conditioned on $\{X_n = -1\}$ are asymptotically independent and the conditioning fades for large values of n , consequently

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(A_n(t) \cap A'_n(t')) &= \mathbb{P}(X_i = x_i \forall i \in [1, m]) \mathbb{P}(X_i = x'_i \forall i \in [1, m']) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(A_n(t)) \mathbb{P}(A'_n(t')). \end{aligned}$$

Finally, using (G.1) and (G.2), we have that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\{v_n \text{ has colour } c \text{ in } T_n\} \cap E_n \cap E'_n) = \mathbb{P}(\mathcal{T}(c) \leq \tau_i^*) \mathbb{P}(\mathcal{T}(c) \leq \tau_i^*).$$

and the result follows.





CHAPITRE

V

FACTORISATIONS PRIMITIVES DU
CYCLE

Contenu du Chapitre

A Introduction	104
B Bijections between plane trees, decreasing and increasing factorisations	107
B.i Labeled trees and plane trees	107
B.ii Link between labeled plane trees and minimal factorisations	109
B.iii Bijection between plane trees and decreasing factorisations	110
B.iv Bijection between plane trees and increasing factorisations	113
C Asymptotic behaviour of random uniform decreasing factorisations	116
C.i Discussing the main result	116
C.ii Lamination processes	117
C.iii Topology	119
C.iv A deterministic result	122
C.v A modification of the cord process	125
C.vi Link between the cords of the decreasing factorisation and the Łukasiewicz path	127



A Introduction

For $n \geq 2$ we define the set of *minimal factorisations* of size n by

$$\mathcal{F}_n := \{(\tau_1, \dots, \tau_{n-1}) : \forall i \tau_i \text{ is a transposition and } \tau_{n-1} \circ \dots \circ \tau_1 = (1 \ 2 \dots n)\}.$$

We also define the set of *decreasing* (resp. *increasing*) factorisations, denoted by \mathcal{F}_n^\downarrow (resp. \mathcal{F}_n^\uparrow), the set of elements in \mathcal{F}_n with the extra condition that, if $\tau_i = (a_i, b_i)$ with $a_i < b_i$, then the a_i 's are in decreasing (resp. increasing) order. For example, $((n \ n-1), (n-1 \ n-2), \dots, (3 \ 2), (2 \ 1))$ is always a decreasing factorisation of $(1 \ 2 \dots n)$; $((8 \ 9), (8 \ 10), (7 \ 8), (2 \ 3), (2 \ 4), (2 \ 5), (1 \ 2), (1 \ 6), (1 \ 7))$ is a decreasing factorisation of $(1 \dots 10)$ and $((1 \ 4), (1 \ 6), (2 \ 4), (3 \ 4), (5 \ 6))$ is an increasing factorisation of $(1 \dots 6)$.

Monotone factorisations are part of a wide family of enumerative problems initiated by Hurwitz [61]. Indeed the study of factorisations of permutations knows many variants and is linked to many other mathematical fields. For example the number of minimal transitive factorisation is counted by the Hurwitz numbers which is linked to ramified covers of the sphere. We suggest to look at [27] and the references therein for further information about this link and for a generalisation of Hurwitz' theorem to general permutations. Another example is minimal factorisations with transposition of the form $(s, s+1)$, called *reduced decompositions*, which have been studied as well [93, 48]. These factorisations are also called *sorting networks* since they are linked to sorting algorithms. Factorisations using star transpositions [87, 63, 59, 50] and cycle of given length [21] have been studied as well. Decreasing factorisations are linked to Jucys-Murphy elements and matrix models [82]. The goal of this article is to study the behaviour of a minimal factorisation chosen uniformly at random in \mathcal{F}_n^\downarrow or \mathcal{F}_n^\uparrow . The asymptotic study of random minimal factorisations has been initiated in [51] and pursued in [52], [98] and [53]. The study of random sorting networks has also been studied in [10] and [9]. The present paper completes the previous articles in the case of random decreasing and increasing factorisations.

Decreasing factorisations (also called *primitive factorisations*) have been studied by Gewurz & Merola [57] who showed that they are counted by Catalan numbers. Two bijections are mentioned in [57]: the first one

$$((a_1 \ b_1), \dots, (a_{n-1} \ b_{n-1})) \rightarrow (a_{n-1}, a_{n-2}, \dots, a_1) \tag{A.1}$$

is a bijection with so-called increasing parking functions (i.e. parking functions $(\pi_1, \dots, \pi_{n-1})$ such that $\pi_1 \leq \dots \leq \pi_{n-1}$), and the second one

$$((a_1 \ b_1), \dots, (a_{n-1} \ b_{n-1})) \rightarrow (b_{n-1}, b_{n-2}, \dots, b_1) \tag{A.2}$$

is a bijection with 231-avoiding permutations. Increasing factorisations appear in [64] and by the generating function in Theorem 5.2, they are also counted by Catalan numbers.

In this work, we explain how both decreasing and increasing factorisations can be put in bijection with plane trees in a rather unified way. Roughly speaking, it is based on the standard coding of general minimal factorisations by labelled trees. More precisely, a minimal factorisation $((a_1 \ b_1), \dots, (a_{n-1} \ b_{n-1}))$ is encoded by the tree with n vertices, labelled from 1 to n , such that an edge labelled i is drawn between the vertices a_i and b_i for all $1 \leq i \leq n-1$. A minimal factorisation can be read on its associated labelled tree, actually, the vertex labels are redundant (see [52, Lemma 2.6]), meaning that the minimal factorisation can be uniquely retrieved from the edge labels. Our bijections state that the edge labels are also redundant if we are restricted to decreasing or increasing factorisations.

Combinatorial applications. Another important merit of our approach is that the bijections allow to obtain information on random uniform decreasing and increasing factorisations by studying the associated random plane tree. For example we show the following results :

Proposition V.A.1

Let $(\tau_1, \dots, \tau_{n-1})$ be a random uniform decreasing factorisation of size n with $\tau_i = (a_i \ b_i)$ and $a_i < b_i$. Then

1. $\{b_1, \dots, b_{n-1}\} = \{2, 3, \dots, n\}$ almost surely.
2. $\mathbb{P}(\#\{a_1, \dots, a_{n-1}\} = k) = \frac{n+1}{n-1} \binom{n-1}{k} \binom{n-1}{k-1} \binom{2n}{n}^{-1}$ for all $k \in \{1, \dots, n-1\}$.
3. The convergence

$$\frac{\#\{a_1, \dots, a_{n-1}\} - n/2}{\sqrt{n}} \xrightarrow{x \rightarrow \infty} \mathcal{N}(0, 1/8)$$

holds in distribution where $\mathcal{N}(0, 1/8)$ designates a centered normal distribution with variance 1/8.

Remarque

The first item of Proposition V.A.1 has already been noticed in [57]. As we said previously, the authors even show that the sequence (b_1, \dots, b_{n-1}) is a 231-avoiding permutation, meaning that, if $i < j < k$, then it never happens that $b_k < b_i < b_j$.

Proposition V.A.2

Let $(\tau_1, \dots, \tau_{n-1})$ be a random uniform increasing factorisation of size n with $\tau_i = (a_i \ b_i)$ and $a_i < b_i$. Then

1. The sets $\{a_1, \dots, a_{n-1}\}$ and $\{b_1, \dots, b_{n-1}\}$ are disjoint and their union is $\{1, 2, 3, \dots, n\}$ almost surely.
2. Items 2 and 3 of Proposition V.A.1 are still satisfied for $\#\{a_1, \dots, a_{n-1}\}$.

Remarque

Actually, the function given by (1) is also a bijection between increasing factorisations and decreasing parking functions. Moreover one can pass from an increasing to a decreasing parking function just by reversing the order of the elements. Therefore item 2 of Proposition V.A.2 can be directly deduced from Proposition V.A.1.

Exploiting bijection (A.1), our work brings some applications to increasing (and decreasing) parking functions. For a parking function $\pi = (\pi_1, \dots, \pi_n)$ of size n and $x \in [0, 1]$, set

$$F_\pi(x) := \frac{1}{n} \#\{i : \pi_i \leq nx\}.$$

The following result will be a consequence of Lemma V.C.15 and bijection (A.1).

Proposition V.A.3

If π is uniformly chosen among the increasing parking functions of size n , then the convergence

$$\left(\sqrt{\frac{n}{2}} (F_\pi(x) - x) \right)_{0 \leq x \leq 1} \xrightarrow{n \rightarrow \infty} e \tag{A.3}$$

holds in distribution for the uniform norm, where e is the Brownian excursion on $[0, 1]$.

Remarque

Convergence (A.3), without the $\sqrt{2}$, is true when π is a uniform minimal factorisation (see [45, Theorem 15]).

Geometrical applications. Our bijections will allow us to study the so-called *lamination process* associated with a random uniform decreasing factorisation. More precisely, let $(\tau_1^n, \dots, \tau_{n-1}^n)$ be a random uniform element of \mathcal{F}_n^\downarrow where $\tau_i^n = (a_i^n, b_i^n)$ with $a_i^n < b_i^n$. The lamination process $(L_k^n)_{1 \leq k \leq n-1}$ is defined by

$$L_k^n := \bigcup_{\ell \leq k} \left[\exp\left(-2i\pi \frac{a_\ell^n}{n}\right), \exp\left(-2i\pi \frac{b_\ell^n}{n}\right) \right].$$

It is a process with values in the set of laminations which are unions of non-crossing cords of the unit disk in the complex plane. The main result of this article, besides the bijections, is the following convergence

Théorème V.A.4

The convergence

$$(L_{[nt]}^n)_{0 \leq t \leq 1} \xrightarrow{n \rightarrow \infty} (L_t(e))_{0 \leq t \leq 1}$$

holds in distribution in the space $\mathbb{D}([0, 1], \mathbb{L})$ of càdlàg functions equipped with the Skorokhod's J_1 topology.

All the topological details and proper definitions can be found in Section C entirely devoted to the proof of Theorem V.A.4. The limiting process $L(e)$ is what we call the *linear Brownian lamination process* and is also properly defined in Section C. The terminology comes from the fact it can be constructed using the normalised Brownian excursion e (which, informally speaking is a Brownian motion conditioned to reach 0 at time 1 and to be positive on $[0, 1]$) by "reading it" from right to left. For $t = 1$, $L_1(e)$ is Aldous' Brownian triangulation [5]. Figure A.1 gives a visual representation of this process. This theorem completes the study of [51] and [98] for random uniform minimal factorisations where the same geometrical point of view has been adopted. Informally speaking, Theorem V.A.4 encodes how "large" transpositions occur when reading, one after the other, the transpositions of a random uniform decreasing factorisation. To keep this introduction short we postpone discussions about Theorem V.A.4 to Section C.i

Plan of the chapter. In Section B we construct a bijection between \mathcal{F}_n^\downarrow (resp. \mathcal{F}_n^\uparrow) and the set of plane trees with n vertices. In Section C we study the asymptotic behaviour of a uniform decreasing factorisation of size

n when n grows to infinity taking advantage of the bijection found in the previous section in order to prove Theorem V.A.4.

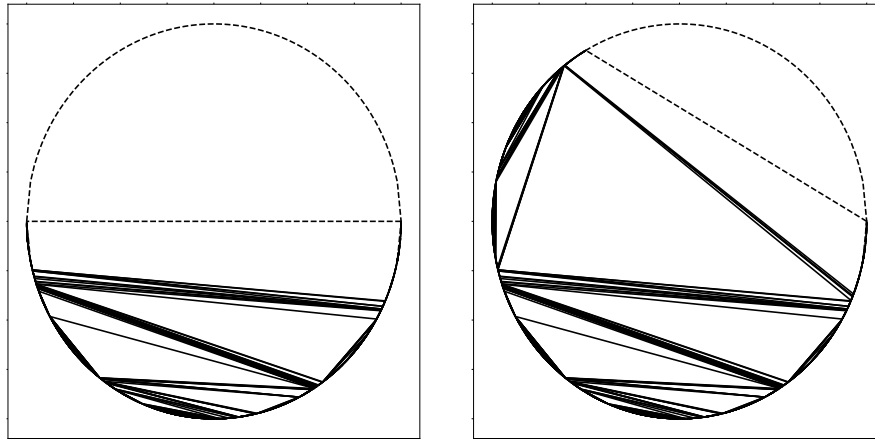


FIGURE A.1 – A representation of $L_{1/2}(e)$ on the left and $L_{2/3}(e)$ on the right.

B Bijections between plane trees, decreasing and increasing factorisations

B.i Labeled trees and plane trees

In this section we recall some basic concepts about trees and introduce some convenient notation that will be used throughout this article.

A *tree* is viewed here as a undirected connected acyclic graph. A tree is said to be *labeled* if vertices, or edges, are associated with some real number, called the label of the vertex or the edge. Note that not all the vertices, or edges, of a labeled tree are required to be labeled. We will sometimes emphasize this fact by saying that the tree is *partially labeled*. We will also use the following terminology : a V-labeled tree is a tree where all the vertices, but none of the edges, are labeled, an E-labeled tree is the other way around and an EV-labeled tree is a tree where all the vertices and edges are labeled. If a is a real number and t a labeled tree, we will denote by $v(a, t)$, or simply $v(a)$ if the underlying tree t is unambiguous, the set of vertices of t labeled a . If $v(a)$ contains only one vertex, we identify this vertex with the singleton $v(a)$.

A *plane tree* is a rooted tree (i.e. a vertex, called the *root*, is distinguished) where, for every vertex v , a total order is given on the edges *stemming from v* (i.e coming from v and going into a child of v). Denote by Π_n the set of plane trees with n vertices. The labels of a labeled plane tree t are said to be *compatible* with the plane order if for every vertex v of t , the natural order of the labels of the edges stemming from v is a restriction of the plane order. Denote by Π'_n the set of EV-labeled plane trees with n vertices where the vertices are labeled from 1 to n (the root being labeled 1), edges are labeled from 1 to $n - 1$ and such that the labels are compatible with the plane order. In all the figures, when representing a plane tree, the edges stemming from a given vertex will be ordered from left to right. Thus, the smallest edge, for the plane order, will be the first to appear on the left and the biggest will be the furthest on the right. Also, the root will be always represented with a double

circle.

If t is a plane tree then we can define a total order between the vertices of t which is the lexicographic order, also called the depth-first search order, when we encode every vertex v with the list of edges coming from the root and leading to v (see Figure B.3 for an example). We denote by \leq this order and the strict version is denoted by $<$. Observe that an ancestor of a vertex v is always smaller than v for the lexicographic order.

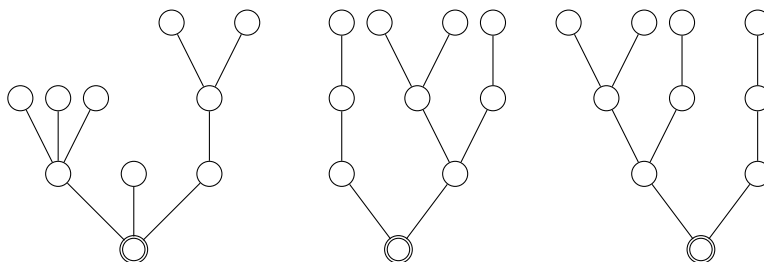


FIGURE B.1 – 3 plane trees with 10 vertices. Notice that the last two do not represent the same plane tree. We denote by t the tree on the left, it will be our example tree for what follows.

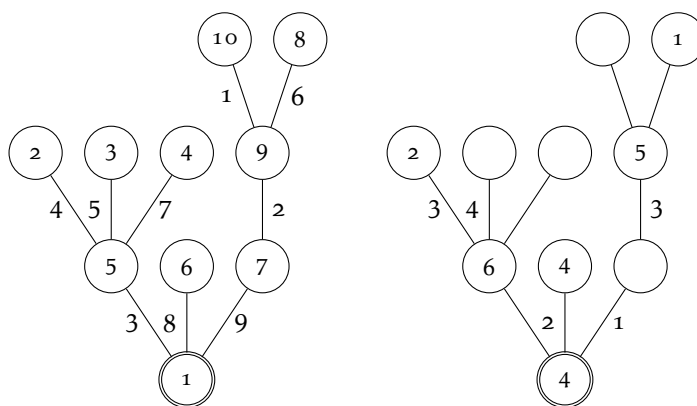


FIGURE B.2 – On the left t has been labeled to get $t' \in \Pi'_{10}$ (notice that the labels of edges stemming from a same vertex are growing from left to right). On the right t has been partially labeled. This labelling is not compatible with the plane order because the edge label 2 is on the left of 1.

We end this section with the definition of the Łukasiewicz path. Let t be a plane tree with n vertices and v_1, \dots, v_n be its vertices ordered in the lexicographic order. The Łukasiewicz path associated with t is the finite sequence $(S_k)_{0 \leq k \leq n}$ such that, for all $0 \leq k \leq n$, $S_k + k$ is the number of edges stemming from a vertex v_ℓ with $\ell \leq k$ (e.g. $S_0 = 0$ and $S_n = -1$). See Figure B.3 for a better visualisation.

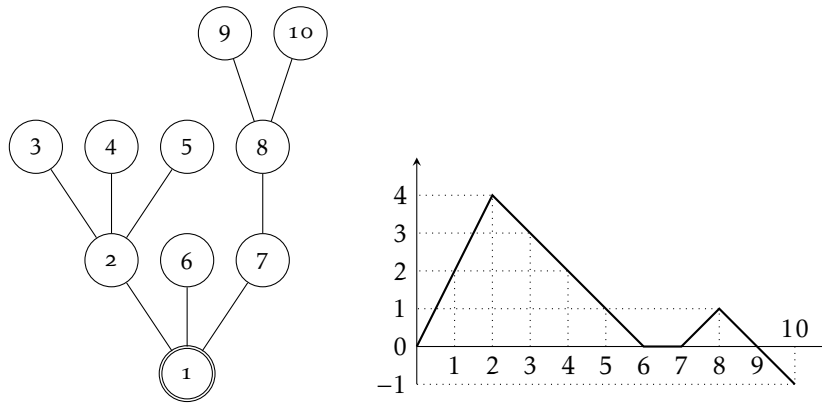


FIGURE B.3 – On the left, the vertices of t has been labeled according to their lexicographic order. On the right is represented the Łukasiewicz path associated with t .

B.ii Link between labeled plane trees and minimal factorisations

Définition V.B.1

Let f be a minimal factorisation of size n . We define three trees $T_1(f)$, $T_2(f)$ and $T_3(f)$ in the following way :

- $T_1(f)$ is the EV-labeled plane tree in Π'_n where an edge labeled k is drawn between the vertices labeled i and j iff $\tau_k = (i, j)$.
- $T_2(f)$ is the E-labeled plane tree obtained from $T_1(f)$ (we keep the shape, the root and the plane order of $T_1(f)$) by removing the vertex-labels but keeping the edge-labels.
- $T_3(f)$ is the (unlabeled) plane tree obtained from $T_1(f)$ by removing the vertex-labels and the edge-labels.

For example, in Figure B.2, $t' = T_1(f)$, where $f = ((9\ 10), (7\ 9), (1\ 5), (2\ 5), (3\ 5), (8\ 9), (4\ 5), (1\ 6), (1\ 7))$ is a minimal factorisation of size 10, and $t = T_3(f)$. A priori, from the definition above, $T_1(f)$ is only a graph. However it is not hard to see that $T_1(f)$ must be connected, and since it has $n - 1$ edges and n vertices, it must be a tree. The map T_1 is clearly an injection from the set of all minimal factorisations of size n and the set Π'_n . However, it is not surjective since the labels of a tree $t \in \Pi'_n$ must satisfy some conditions and can't be chosen freely in order to be in the image of T_1 . Actually, the latter conditions on the labels are such that, if we erase the vertex-labels of a tree $T_1(f)$, then it is possible to retrieve them. In other words, the map T_2 is also an injection. Moreover, any E-labeled plane tree with $n - 1$ edges compatibly labeled from 1 to $n - 1$ can be uniquely labeled to get an element in the image of T_1 . We will describe, right after this discussion, an algorithm (see [52, Section 2.3]) that performs this labeling. To summarize, T_2 is a bijection from the set of minimal factorisations of size n to the set of E-labeled plane trees with $n - 1$ edges compatibly labeled from 1 to $n - 1$. In other words, for minimal factorisations, the vertex-labels are redundant. The goal of the next two sections will be to show that adding the decreasing, or increasing, condition also makes the edge-labels redundant. In other words we will show that T_3 is a bijection from the set of decreasing (resp. increasing) factorisations of size n and the set Π_n of (unlabeled) plane trees with n vertices.

Let t be a E-labeled plane tree with $n - 1$ edges compatibly labeled from 1 to $n - 1$. In what follows we describe an "exploration" algorithm Next that takes the tree t and an integer $k \in \{0, 1, \dots, n - 1\}$ as arguments and gives the label $k + 1$ to a vertex of t , such that, if we apply successively $\text{Next}(t, 0), \text{Next}(t, 1), \dots, \text{Next}(t, n - 1)$ then t gets the unique labeling that includes him in the image of T_1 . First $\text{Next}(t, 0)$ gives the label 1 to the root of t . For $k \geq 1$, the algorithm starts from the vertex $v_0 = v(k)$ of t labeled k . Then it follows the longest possible path possible of edges e_1, \dots, e_ℓ such that :

- e_1, \dots, e_ℓ is a path of edges meaning that for all $0 < i < \ell$, e_i and e_{i+1} share a common vertex v_i .
- e_1 is the edge with smallest label adjacent to v_0 .
- For all $0 < i < \ell$, e_{i+1} is the successor of e_i meaning that e_{i+1} has the smallest label among the edges of v_i having a label greater than e_i 's label.

Starting from v_0 and following the path e_1, \dots, e_ℓ leads to the vertex of t that will receive the label $k + 1$. For instance, if we take the tree t' in Figure B.2 again, then $\text{Next}(t', 1)$ goes through the edges labeled 3 and 4 to end up on the vertex labeled 2. We can also check that $\text{Next}(t', 6)$ passes through the edges labeled 8 and 9 to end up on the vertex labeled 7.

Observe that when applying $\text{Next}(t, 0), \text{Next}(t, 1), \dots, \text{Next}(t, n - 1)$ successively, then, after entering the first time in a fringe subtree of t (i.e. a subtree of t formed by one vertex of t and all his descendants), the exploration exits this fringe subtree only after labelling all its vertices (see [52, Proof of Lemma 2.10]). This remark leads to the following fact :

Lemme V.B.2

Let f be a minimal factorisation and let $T_1(f)$ be its associated EV-labeled plane tree. For every pair of vertices $u < v$ (i.e. u is before v in the lexicographic order) such that u is not an ancestor of v , the label of u is smaller than the label of v in $T_1(f)$.

Nevertheless, if u is an ancestor of v then the label of u could be greater than the label of v . This lemma will be useful in the proof of Proposition V.B.5.

B.iii Bijection between plane trees and decreasing factorisations

The aim of this section is to show that the map T_3 , defined in Section B.ii, is a bijection between the set \mathcal{F}_n^\downarrow of decreasing factorisations of size n and the set Π_n of plane trees with n vertices. The result is presented in Proposition V.B.3, the proof gives an explicit construction of the inverse function of T_3 .

Proposition V.B.3

The map T_3 is a bijection between \mathcal{F}_n^\downarrow and Π_n .

The set of plane trees Π_n is known to be of cardinality the n -th Catalan's number. So, Proposition V.B.3, allows us to deduce the cardinality of \mathcal{F}_n^\downarrow . The cardinality of \mathcal{F}_n^\downarrow was already known using bijective (see [57])

and non-bijective (see [64]) proofs, so the result is not new, but it gives a new bijective proof which will be relevant for studying its probabilistic properties.

Corollaire V.B.4

The cardinality of \mathcal{F}_n^\downarrow is $|\mathcal{F}_n^\downarrow| = \frac{(2n)!}{n!(n+1)!}$.

Preuve de la Proposition V.B.3

We shall construct explicitly the inverse bijection of T_3 , i.e. starting from a rooted plane tree we shall explain how to label its vertices and edges so that it corresponds by T_3 to the EV-labeled plane tree of a unique decreasing factorisation. Let t, t' be two partially labeled plane trees, we write $t \sim t'$ if t and t' represent the same plane tree (when we forget all the labels) and for every vertex, or edge, w , if w is labeled in both t and t' , then it has the same label in t and t' . Beware that \sim is not transitive so it's not an equivalence relation. Recall that the map T_1 is defined in the beginning of Section B.ii. The proof just boils down to showing that for every $t \in \Pi_n$ there exists a unique $f \in \mathcal{F}_n^\downarrow$ such that $t \sim T_1(f)$.

Let v_1, \dots, v_n be the vertices of t ordered in the lexicographic order (so v_1 is the root of t) and Let S be the Łukasiewicz path associated with t . Recall that the Łukasiewicz path is defined in Section B.i. For all $k \in \{1, \dots, n\}$ we denote by t_k the partially labeled plane tree obtained from t where :

- $\forall i \leq k, v_i$ gets the label i .
- $\forall i \leq k$, the edges stemming from v_i get the labels from $n-1-(S_i+i)$ to $n-1-(S_{i-1}+i-1)$.

In other words, one explores the vertices of t in lexicographic order and labels their outgoing edges in increasing order from left to right using the largest possible labels. We also write $t_0 = t$ (see Figure B.4 for an example).

Let f be a decreasing factorisation. We will prove by induction that if f is such that $t \sim T_1(f)$, then for all $0 \leq k \leq n$, $t_k \sim T_1(f)$.

It is true for $k = 0$. Let's assume it is true for a $k \in \{0, \dots, n-1\}$. We advise to look at Figure B.4 while reading the proof to have a better visualisation.

Let e_1, \dots, e_ℓ (with $\ell > 0$) be the shortest path of edges starting from v_k and going to v_{k+1} . Then, let $u_1 = v_k, u_2, \dots, u_{\ell+1} = v_{k+1}$ be the successive vertices the path e_1, \dots, e_ℓ goes through. Since v_{k+1} is the successor of v_k for the lexicographic order in t , one can easily check that :

- u_1, u_2, \dots, u_ℓ are the successive ancestors of v_k up to u_ℓ in t and v_{k+1} is a child of u_ℓ .
- if v_k is not a leaf, $\ell = 1$ and e_1 is the smallest edge (for the plane order) among the edges stemming from v_k in t .
- if v_k is a leaf, $\ell > 1$ and e_ℓ is the smallest edge among the edges stemming from u_ℓ that are greater than $e_{\ell-1}$ in t .
- for all $0 < i < \ell - 1$, e_i is the greatest edge among the edges stemming from u_{i+1} in t .

Now we observe that if v is a non-root vertex of t , then the labels, in t_n , of the edges stemming from v are all smaller than the label of the edge between v and its parent. Using this observation, what precedes

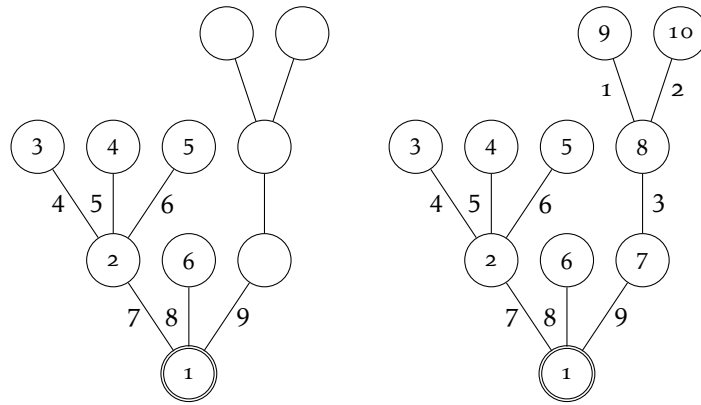


FIGURE B.4 – Illustration of the proof of Proposition V.B.3. The partially labeled plane tree t_6 is on the left and the EV-labeled plane tree t_{10} is on the right.

and by induction we get that e_1, \dots, e_ℓ is exactly the path that $\text{Next}(T_1(f), k)$ would take, so v_{k+1} must have the label $k+1$ in $T_1(f)$. It remains to check that the labels of the edges stemming from v_{k+1} in $T_1(f)$ must be the largest among the unclaimed edge-labels in t_{k+1} . Let e be an edge between vertices u and w such that e is not labeled yet in t_{k+1} . Denote by a and b the labels of u and w in $T_1(f)$. By construction of t_{k+1} , $v_{k+1} < u, w$ so $k+1 < a, b$. Since f is a decreasing factorisation, the label of e in $T_1(f)$ must be smaller than the label of the edges stemming from v_{k+1} . This implies the initial statement and ends the proof of $t_{k+1} \sim T_1(f)$ which concludes the induction.

Finally we have $T_1(f) = t_n$. By injectivity of T_1 , uniqueness is shown. It remains to prove existence. Since T_2 is surjective, we can find f a minimal factorisation such that $T_2(f)$ is the tree t_n where the vertex-labels are erased. First, notice that, applying $\text{Next}(T_2(f), 0), \dots, \text{Next}(T_2(f), n-1)$ gives $T_1(f) = t_n$. This can be easily seen if we follow the same scheme as for uniqueness and see that the algorithm Next always labels the successor for the lexicographic order in $T_2(f)$. From $T_1(f)$ we easily check that f is a decreasing factorisation.




The bijection T_3 and its inverse function will be useful to prove some asymptotic results in Section C for a decreasing factorisation chosen uniformly at random. The bijection has also direct non-asymptotic consequences presented in Proposition V.A.1 which we prove below.

Preuve de la Proposition V.A.1

Let $t := T_1(\tau_1, \dots, \tau_{n-1})$. For 1, notice that an integer i belongs to $\{b_1, \dots, b_{n-1}\}$ if and only if $v(i, t)$ is adjacent to a vertex whose label is smaller than i . Recall that, in t , the vertices are labeled according to their lexicographic order. So every vertex $v(i, t)$, for $i > 1$, has a parent whose label is smaller than i . Consequently $\{b_1, \dots, b_{n-1}\} = \{2, 3, \dots, n\}$.

For 2, notice that an integer i belongs to $\{a_1, \dots, a_{n-1}\}$ if and only if $v(i, t)$ is adjacent to a vertex whose label is greater than i . This is the case exactly when $v(i, t)$ is not a leaf. Therefore, the cardinality of $\{a_1, \dots, a_{n-1}\}$ is distributed like the number of non-leaf vertices in a uniform plane tree with n vertices.

This distribution is known [42] and is given by Narayana’s numbers.

For 3, we apply [69, Example 2.1] which gives a central limit theorem for the number of leaves of a critical Bienaymé-Galton-Watson tree with n vertices and reproduction law ξ . Taking ξ to be a geometrical law of parameter $1/2$, gives the law of a uniform plane tree. 

B.iv Bijection between plane trees and increasing factorisations

The aim of this section is to show that T_3 , defined in Section B.ii, is a bijection between the set \mathcal{F}_n^\uparrow of increasing factorisations of size n and the set Π_n of plane trees with n vertices, following the same principle as in Section B.iii. The result is presented in Proposition V.B.5 and the proof gives an explicit construction of a bijection.

Proposition V.B.5

The map T_3 is a bijection between \mathcal{F}_n^\uparrow and Π_n .

Corollaire V.B.6

The cardinality of \mathcal{F}_n^\uparrow is $|\mathcal{F}_n^\uparrow| = \frac{(2n)!}{n!(n+1)!}$.

As for the decreasing case, there are non-bijective proofs of Corollary V.B.6 (see [64]) so this result is not new.

Preuve de la Proposition V.B.5

If t is a partially labeled plane tree and v is a vertex of t , then we denote by $\text{sub}(v)$ the fringe subtree of t , composed of v and all its descendants, and $|\text{sub}(v)|$ the number of vertices of that subtree (e.g. if v is a leaf of t , then $|\text{sub}(v)| = 1$). Similarly to the proof of Proposition V.B.3, the proof boils down to show that for every $t \in \Pi_n$ there exists a unique $f \in \mathcal{F}_n^\uparrow$ such that $t \sim T_1(f)$.

To show the latter claim, we will, once again, construct by induction a sequence of partially labeled plane trees t_1, \dots, t_m with $t_0 = t$ and $t_m \in \Pi'_n$ by gradually adding labels on t . As before, denote by v_1, \dots, v_n the vertices of t ordered in the lexicographic order. We also denote by ν_k (resp. ε_k) the smallest unclaimed positive integer vertex (resp. edge) label of t_k (e.g. $\nu_0 = \varepsilon_0 = 1$ and $\nu_m - 1 = \varepsilon_m = n$). We say that a vertex of t is of *type 1* if it is a leaf or if its height is an even number (e.g. the root is of type 1 because its height is 0). The other vertices are said to be of *type 2*. To get t_1 we simply label the root v_1 with 1 and the edges stemming from v_1 with $1, 2, \dots, \varepsilon_1 - 1$ where $\varepsilon_1 - 1$ is the degree of v_1 . Now we explain how to construct t_{k+1} from t_k . Denote by v the vertex that has been assigned label ν_{k-1} in t_k . If $v = v_n$ then the procedure ends and we set $m = k$. Otherwise, let u' be the successor of v for the lexicographic order and let u be the parent of u' . The tree t_{k+1} is constructed in the following way (look at Figures B.5 and B.6):



- Case 1 : If u' is a leaf. We give the label v_k to u' and if the edge between u and u' has no label in t_k we label it by ε_k .
- Case 2 : If u' is not a leaf and u is of type 2. We give the label v_k to u' , we give the labels $\varepsilon_k, \dots, \varepsilon_{k+j}-1$ to the edges e_1, \dots, e_j stemming from u' and the label $\varepsilon_k + j$ to the edge e_{j+1} between u' and u .
- Case 3 : If u' is not a leaf and u is of type 1. Denote by u'' the successor of u' for the lexicographic order. We give the label $v_k + |\text{sub}(u')| - 1$ to u' and label v_k to u'' . We give the labels $\varepsilon_k, \dots, \varepsilon_{k+i} - 1$ to the edges e_1, \dots, e_i stemming from u'' and the label $\varepsilon_k + i$ to the edge e_{i+1} between u'' and u' .

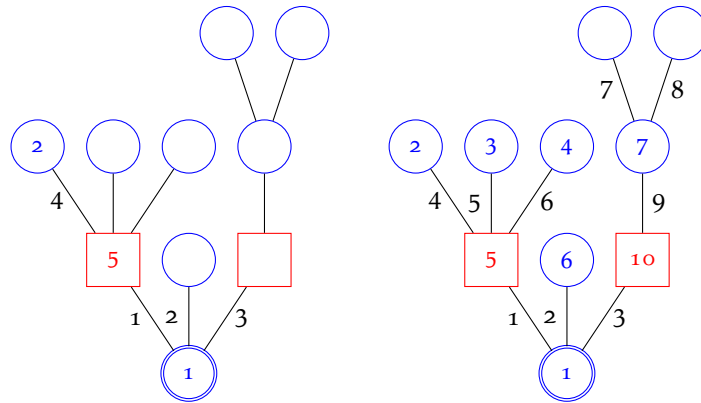


FIGURE B.5 – Illustration of the proof of Proposition V.B.5. The partially labeled plane tree t_2 is on the left and t_6 is on the right. Blue circles (resp. red squares) represent vertices of type 1 (resp. 2).

We can make useful observations about the construction of the t_k 's.

- Observation 1 : For every k , denote by i_k the number of labeled vertices in t_k . The labeled vertices in t_k are exactly the vertices v_1, \dots, v_{i_k} . Notice that $i_{k+1} - i_k = 1$ or 2 so in particular $m \leq n$. For every $w, w' \leq v_{i_k}$ that share a common edge e , e is labeled in t_k . More precisely, the labeled edges in t_k are exactly the edges adjacent to vertices $w \leq v_{i_k}$ of type 1.
- Observation 2 : A vertex w is of type 1 iff there exists k such that w has label v_{k-1} in t_k . And in this case $w = v_{i_k}$.
- Observation 3 : Let $w = v_i$ be a vertex distinct from the root (so $i > 1$) or from a leaf with $i \leq i_k$. Denote by e_1, \dots, e_j all the edges stemming from w in increasing order that are labeled in t_k ($j \geq 1$ by observation 1). Also denote by e the edge linking w to its parent (it is labeled in t_k by observation 1). If w is of type 1 then $j = \text{deg}(w) - 1$ and e 's label is greater than e_j 's label. If w is of type 2 then e 's label is smaller than e_1 's label.
- Observation 4 : Let w, w' be two vertices such that w is of type 2 and w' is a descendant of w . If w and w' are both labeled in t_k (namely, if $w' \leq v_{i_k}$) then the label of w' is smaller than the label of w .

Now, if $f \in \mathcal{F}_n^\uparrow$ is such that $t \sim T_1(f)$ then we claim that for every k , $t_k \sim T_1(f)$. It is clearly true for $k = 0$ since $t_0 = t$. It is also true for $k = 1$ since f is increasing, implying that the transpositions containing 1 must appear first in f . Now assume it is true for a certain $k < m$. We use the same notations and the same cases used in the construction of the t_k 's. Consider the procedure P which consists in applying the algorithm Next several times on $T_1(f)$ starting from v (which has label v_{k-1} in $T_1(f)$) up to the vertex of label v_k in $T_1(f)$. Thus, the procedure P consists in applying

$\text{Next}(T_1(f), v_{k-1}), \text{Next}(T_1(f), v_{k-1} + 1), \dots, \text{Next}(T_1(f), v_k - 1)$. We claim that this procedure goes through the vertex u' (but it doesn't mean it stops on u'). To see this, we need to study two cases. If v is not a leaf (so $u = v$), since v is of type 1 (observation 2) by using observation 3 we conclude that P starts by going up the edge between v and u' . If v is a leaf, obviously P has no choice but to take the edge between v and its parent. Then observation 3 implies that P takes the path down to u (stopping at every vertex of type 2) and then goes to u' . The question is now to see if the procedure P stops on u' or not.

- Case 1 : If u' is a leaf, the procedure stops on u' which imply that u' has label v_k in $T_1(f)$. Notice that if u is of type 1, then the edge between u and u' is already labeled in t_k (observation 3) so there is nothing left to prove in order to conclude $t_{k+1} \sim T_1(f)$. Suppose that u is of type 2, we want to show that the edge between u and u' has label ε_k in $T_1(f)$. Let e be an edge between vertices $v(a)$ and $v(b)$, such that e is not yet labeled in t_{k+1} . Suppose that $v(a) < v(b)$. Observation 1 implies that, either $u' < v(a)$, either $u' < v(b)$ and $v(a)$ is an ancestor of u' of type 2. Using observation 4 and Lemma V.B.2, in both cases a and b are greater than the label of u' in $T_1(f)$, namely v_k . Since f is an increasing factorisation, it implies that the label of e in $T_1(f)$ must be greater than the label between u and u' . It shows that $t_{k+1} \sim T_1(f)$.
- Case 2 : In this case, P stops on u' . Indeed, because u is of type 2 it has a greater label than all its descendants (in particular u') in $T_1(f)$ (this comes from the remark preceding Lemma V.B.2 combined with observation 3 : applying successively Next to $T_2(f)$ will label u after all its descendants). By contradiction, if P doesn't stop on u' then, u' has a greater label than its first child u'' in $T_1(f)$ (by the remark preceding Lemma V.B.2). It also implies that the edge between u and u' is smaller than the edge between u' and u'' . This contradicts the fact that f is increasing. In conclusion P stops on u' , thus u' has label v_k in $T_1(f)$. It remains to show that the edges adjacent from u' are labeled in $T_1(f)$ as in t_{k+1} . Let e be an edge between two vertices labeled $v(a)$ and $v(b)$ such that e is not yet labeled in t_{k+1} . Suppose that $v(a) < v(b)$. If u' is an ancestor of $v(a)$, since P stops on u' , u' has a label smaller than a and b . If $v(a)$ is an ancestor of u' then $v(a)$ is of type 2 and, like in case 1, we still have that a, b are greater than the label of u' . Otherwise, by Lemma V.B.2, the inequality still holds. Since f is increasing the label of e in $T_1(f)$ must be greater than the labels of the edges adjacent to u' which shows that $t_{k+1} \sim T_1(f)$.
- Case 3 : This time P doesn't stop at u' . We can show this by contradiction, like for case 2. If P stops at u' then the edge between u and u' has a greater label than all the edges stemming from u' . Moreover it implies $\text{label}(u) < \text{label}(u') < \text{label}(u'')$. Thus the procedure P won't stop at u' . This shows that u' doesn't have the label v_k on $T_1(f)$ and actually has the label $v_k + |\text{sub}(u')| - 1$ (because it is the last vertex seen by Next in $\text{sub}(u')$). From this point we can follow the same reasoning as for case 2 and show that $t_{k+1} \sim T_1(f)$.

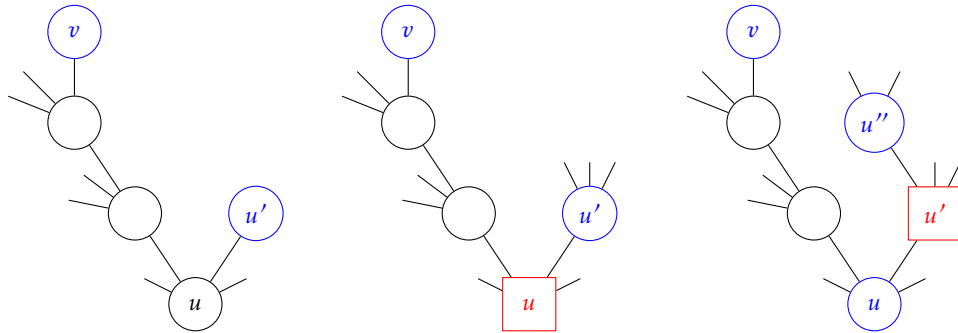


FIGURE B.6 – Illustration of the three cases in the proof of Proposition V.B.5 when v is a leaf. Case 1 is illustrated on the left, case 2 in the middle and case 3 on the right. Black circles represent indifferently vertices of type 1 or 2.

In conclusion we have $t_k \sim T_1(f)$ for all k , so $t_m = T_1(f)$. By injectivity of T_1 , uniqueness is shown. It remains to prove existence. Since T_2 is surjective, we can find f a minimal factorisation such that $T_2(f)$ is the tree t_m where the vertex-labels are erased. First, notice that, applying $\text{Next}(T_2(f), 0), \dots, \text{Next}(T_2(f), n-1)$ gives $T_1(f) = t_m$. This can be easily seen if we follow the same scheme as for uniqueness and see that the algorithm Next labels the vertices of $T_2(f)$ as described above. From $T_1(f)$ we easily check that f is an increasing factorisation.



Now we can prove the result of Proposition V.A.2 presented in the introduction.

Preuve de la Proposition V.A.2

Let $t := T_1(\tau_1, \dots, \tau_{n-1})$. Notice that $v(a_1, t), \dots, v(a_{n-1}, t)$ are exactly the vertices in t at even height (the root being at height 0). Similarly $v(b_1, t), \dots, v(b_{n-1}, t)$ are exactly the vertices in t at odd height. This implies the first point of the corollary. For the second point we use the fact that the number of vertices at even height is distributed like the number of leaves (see [43]). Finally we conclude using the same arguments developed in the proof of Proposition V.A.1.



C Asymptotic behaviour of random uniform decreasing factorisations

C.i Discussing the main result

We begin Section C with a discussion on Theorem V.A.4. Theorem V.A.4 is an analog to Theorem 1.2 of [98] where the author studies the lamination process associated to a random uniform minimal factorisation, whereas our theorem concerns random uniform decreasing factorisations. Both theorems give a convergence in the sense of Skorokhod’s J_1 topology. In the case of minimal factorisations, the convergence of the whole lamination process is a generalisation of the fixed time convergence shown in [51]. In Theorem 1.2 of [98], the right scaling factor is \sqrt{n} meaning that macroscopic cords appear after a time of order \sqrt{n} , whereas in Theorem V.A.4, the right scaling factor is n . Another difference with the minimal case is that, in the decreasing case, conditionally on ϵ , the lamination process $L(\epsilon)$ is deterministic while in the minimal case, there is a

second layer of randomness. Indeed, conditionally on ϵ , the cords of the limit process are chosen randomly by throwing points under the curve of ϵ in a Poissonian way.

Theorem V.A.4 implies that for any continuous functional $F : \mathbb{D}([0, 1], \mathbb{L}) \rightarrow E$, where E is a metric space, $F(L^n)$ converges towards $F(L(\epsilon))$ in distribution. For instance the functional giving the size of the current longest cord is continuous for the J_1 topology (in this case E is the set of càdlàg functions from $[0, 1]$ to \mathbb{R}).

We think that there is no result equivalent to Theorem V.A.4 in the case of increasing factorisations. More precisely, if $(L^n_{[nt]})_{0 \leq t \leq 1}$ is the lamination process associated with the random uniform increasing factorisation $(\tau_1^n, \dots, \tau_{n-1}^n)$, we believe that there is no convergence in distribution of this lamination process in the sense of Skorokhod’s J_1 topology. Indeed, using the bijection of Section B.iv, one can see that new macroscopic cords could appear at times arbitrarily close in L^n , for n large enough, which would make the J_1 convergence impossible. More precisely, those macroscopic cords would appear at every branching point of the associated plane tree giving birth to, at least, two subtrees with a non negligible mass of vertices.

To prove Theorem V.A.4 we introduce, in the next sections, some random excursion f_n associated to the decreasing factorisation $(\tau_1^n, \dots, \tau_{n-1}^n)$. There are then two main steps to prove the theorem dealt in the next sections. The first one consists in showing that $L(f_n)$ converges in distribution towards $L(\epsilon)$ for the Skorokhod distance d_S . The second step consists in showing that the quasi-distance d'_S (introduced in Section C.iii) between L^n and $L(f_n)$ goes to 0 in probability when n tends to infinity. Finally, combining those two steps and using Lemma V.C.8 implies the desired result.

C.ii Lamination processes

The aim of this section is to properly introduce the lamination process associated with a random uniform decreasing factorisation of size n (which is quickly defined in the introduction). The ultimate goal being to study the asymptotic behaviour of this process when n tends to infinity. Our main result is Theorem V.A.4 stating that the lamination process converges in distribution. To make sense of this convergence, first, we need to describe the spaces we are working in and, second, to define the topology we are using. Consider the *unit disk* $D := \{z \in \mathbb{C} : |z| \leq 1\}$ of the complex plane.

Définition V.C.1 : Cord

Let $u, v \in [0, 1]$ with $u \leq v$. The *cord* with end points u and v , denoted by $[[u, v]]$, is the subset of the unit disk formed by the segment $[e^{-2i\pi u}, e^{-2i\pi v}]$. More precisely,

$$[[u, v]] := \{te^{-2i\pi u} + (1-t)e^{-2i\pi v} : t \in [0, 1]\}.$$

A cord $[[u, v]]$ is said to be *trivial* if $u = v$ or if $0 = u = 1 - v$. Two cords $[[u, v]]$ and $[[a, b]]$ are said to be *non-crossing* if they do not intersect in the interior of the unit disk, in other words, if $[[u, v]] \cap [[a, b]] \subset \{u, v\}$.

Définition V.C.2 : Lamination

A *lamination* is a non-empty subset $L \subset D$ of the unit disk such that :

- (i) L is compact.
- (ii) L can be written as a union of non-crossing cords.

The set of all laminations is denoted by \mathbb{L} .

Remarque

The union in (ii) has no particular requirements, it is not necessarily finite or countable. A finite union of non-crossing cords is always a lamination. It is not hard to see that a subset of the unit disk is a lamination if and only if it is the closure of a union of non-crossing cords.

We can now introduce the object of interest. Let $(\tau_1^n, \dots, \tau_{n-1}^n)$ be a random uniform decreasing factorisation of the cycle $(1 \dots n)$. For all i , write $(a_i^n, b_i^n) = \tau_i^n$ with $a_i^n < b_i^n$.

Définition V.C.3 : Discrete lamination process

For $1 \leq k \leq n - 1$, we define :

$$L_k^n := \bigcup_{i \leq k} \left[\left[\frac{a_i^n}{n}, \frac{b_i^n}{n} \right] \right].$$

We also set $L_0^n := [[0, 0]] = \{(1, 0)\}$ and $L_n^n := L_{n-1}^n$. The process $L^n := (L_{[nt]}^n)_{0 \leq t \leq 1}$ is called the *lamination process associated with $(\tau_1^n, \dots, \tau_{n-1}^n)$* or, in short, the *discrete lamination process*, to emphasise the fact that L^n is a finite union of cords appearing at discrete times.

The denomination "lamination" is justified for L^n since for all k , the cords of L_k^n do not cross (see [58]). The overall goal of Section C is to show that the process L^n converges in distribution. Below, we describe precisely the limit of L^n , which we call the linear Brownian lamination process.

Denote by \mathcal{E} the set of *excursions*, namely, the set of non-negative continuous functions g on $[0, 1]$ such that $g(0) = g(1) = 0$ and $g(t) > 0$ for all $t \in (0, 1)$. We say that $[[u, v]]$ is a *cord* of $g \in \mathcal{E}$ if $g(u) = g(v) = \min_{[u, v]} g$. The Brownian excursion, denoted by e , is a random element of \mathcal{E} and will be of particular interest to us. Informally the Brownian excursion is a Brownian motion conditioned to be an element of \mathcal{E} . One can construct e by renormalising the excursion of a Brownian motion above, or below, 0 around time 1 (see e.g. [65]).

Définition V.C.4 : Linear Brownian lamination process

For $g \in \mathcal{E}$ and $t \in [0, 1]$ we define :

$$L_t(g) := \bigcup_{\substack{[[u, v]] \text{ cord of } g \\ u \geq 1-t}} [[u, v]].$$

We write $L(g) := (L_t(g))_{0 \leq t \leq 1}$. When $g = e$, $L(e)$ is called the *linear Brownian lamination process*.

The limit of L^n is the linear Brownian lamination process $L(\epsilon)$. Proposition V.C.6 justifies the denomination "lamination" for $L(\epsilon)$ by proving that the cords of ϵ do not cross and form a compact set. Actually this property is satisfied for any excursion with unique local minima. Formally, an excursion g is said to have *unique local minima* if, whenever g reaches two local minima at distinct times s and t , then $g(s) \neq g(t)$.

C.iii Topology

In this section we introduce all the topologies and their associated tools to state formally the convergence of Theorem V.A.4 and prove it. First we need a topology on the set of laminations \mathbb{L} . Since laminations are, by definition, non-empty compact subsets of the unit disk D , it is natural to endow \mathbb{L} with the Hausdorff distance. Here, we recall the definition of the Hausdorff distance in the general setting.

Définition V.C.5 : Hausdorff distance

Let (E, d) be a metric space and denote by K the set of all non-empty, closed and bounded subsets of E . The *Hausdorff distance* d_H is a distance on K given by the following : for all $A, B \in K$

$$d_H(A, B) := \min\{\epsilon \geq 0 : A \subset B^\epsilon \text{ and } B \subset A^\epsilon\}$$

where $U^\epsilon := \{x \in E : d(x, U) \leq \epsilon\}$ is the set of points at distance at most ϵ from U .

In our case we consider the Hausdorff distance on the set of non-empty compact subsets of D , denoted by \mathbb{K} . In this setting, (\mathbb{K}, d_H) is a compact metric space and \mathbb{L} is closed so (\mathbb{L}, d_H) is also compact. Now that \mathbb{L} is endowed with a metric, we can consider càdlàg functions with values in \mathbb{L} . Recall that $f : [0, 1] \rightarrow (E, d)$ is a càdlàg function, where (E, d) is a metric space, if for every $t \in [0, 1]$, f is continuous on the right of t and for every $t \in (0, 1]$, f has a left limit at t . The discrete lamination process L^n is obviously càdlàg (its jumps occur at the times i/n for $1 \leq i \leq n-1$). The following proposition shows that the linear Brownian lamination process is a càdlàg function with values in \mathbb{L} .

Proposition V.C.6

If $g \in \mathcal{E}$ then $L(g) : [0, 1] \rightarrow (\mathbb{K}, d_H)$ is càdlàg on $[0, 1]$ and continuous in 1. Moreover if g has unique local minima then $L_1(g)$ is a lamination (and so is $L_t(g)$ for all $t \in [0, 1]$).

Before proving Proposition V.C.6 we give a simple notation which will be useful for most of the following proves : we say that two cords c and c' are ϵ -close if $d_H(c, c') \leq 2\pi\epsilon$. Notice that the cords $[[u, v]]$ and $[[u', v']]$ are ϵ -close if $|u - u'| \leq \epsilon$ and $|v - v'| \leq \epsilon$.

Preuve de la Proposition V.C.6

- For $t \in [0, 1]$, $L_t(g)$ is compact. Let $(x_n)_n$ be a sequence of $L_t(g)$ that converges to x belonging to the unit disk. For all n , denote by $[[u_n, v_n]]$ a cord of g containing x_n with $1 - t \leq u_n$. By compactity suppose that $(u_n, v_n)_n$ converges towards (u, v) . By continuity $[[u, v]]$ is a cord of g and $1 - t \leq u$. Moreover, we can see that $([[u_n, v_n]])_n$ converges to $[[u, v]]$ for the Hausdorff distance which implies

that x belongs to the cord $[[u, v]]$.

- $L(g)$ is right-continuous. Indeed, by contradiction, assume that we can find $\varepsilon > 0$ and a sequence $([[u_n, v_n]])_n$ of cords of g such that $(u_n)_n$ increases and converges towards u and for every cord $[[u', v']]$ of g with $u \leq u'$, for all n , $[[u_n, v_n]]$ and $[[u', v']]$ are not ε -close. By compactity suppose that $(v_n)_n$ converges towards v . Then by continuity, $[[u, v]]$ is a cord of g and $d_H([[u_n, v_n]], [[u, v]]) \rightarrow 0$ when $n \rightarrow \infty$ which yields a contradiction.
- $L(g)$ has left limits. For all $t \in (0, 1]$, we show that the left limit at t is $K_t := \overline{\cup_{s < t} L_s(g)}$. Let $\varepsilon > 0$ and c_1, \dots, c_n be cords of K_t such that for any cord c of K_t , there is an i such that c and c_i are ε -close. For all i we can find $s_i < t$ such that there is a cord c'_i of $L_{s_i}(g)$ which is ε -close of c_i . Thus for $t > s \geq \max\{s_1, \dots, s_n\}$, $d_H(K_t, L_s(g)) \leq 4\varepsilon$.
- $L(g)$ is continuous in 1. By the last point, the left limit at time 1 is $K_1 = \overline{\cup_{s < 1} L_s(g)}$. Since the cord added at time 1 is the trivial cord $[[0, 1]]$ which also corresponds to the cord added at time 0, $[[1, 1]]$, we have that $\cup_{s < 1} L_s(g) = L_1(g)$. By the first point we know that $L_1(g)$ is closed, so $K_1 = L_1(g)$ which shows the continuity at 1.
- If g has local minima then the cords of $L_1(g)$ do not cross. Indeed if $[[u_1, v_1]]$ and $[[u_2, v_2]]$ are two cords of g with $u_1 \leq u_2$, then it is clear that $u_1 \leq u_2 \leq v_2 \leq v_1$ or $u_1 \leq v_1 \leq u_2 \leq v_2$.



The last task to complete, in order to make the statement of Theorem V.A.4 fully rigorous, is to define the topology on the set of càdlàg functions from $[0, 1]$ to (E, d) denoted by $\mathbb{D}([0, 1], E)$. This topology is of course the celebrated Skorokhod's J_1 topology (see e.g. [66] chapter VI).

Définition V.C.7 : Skorokhod distance

Let $f, g : [0, 1] \rightarrow (E, d)$ be two càdlàg functions with values in a metric space (E, d) . The Skorokhod distance between f and g is

$$d_S(f, g) := \inf_{\varphi \in \Lambda} \max \left\{ \sup_{[0, 1]} |\varphi - Id|, \sup_{[0, 1]} d(f, g \circ \varphi) \right\}$$

where $\Lambda := \{\varphi : [0, 1] \rightarrow [0, 1] : \varphi \text{ is continuous, strictly increasing and bijective}\}$ and $Id \in \Lambda$ is the identity function.

The function d_S defined as above is a distance and induces the so called Skorokhod's J_1 topology on the set of càdlàg functions. This topology is often used to study the convergence of random processes with càdlàg trajectories. At this point, all the tools have been properly defined to fully understand the statement of Theorem V.A.4. From here, up to the end of this section, we give some properties of Skorokhod's J_1 topology which will be useful to prove Theorem V.A.4 but are not required to understand the result. Let

$$\Lambda' := \{\varphi : [0, 1] \rightarrow [0, 1] : \varphi \text{ is continuous, non-decreasing and surjective}\}.$$



Similarly to the Skorokhod distance we define the functions

$$d'_S(f, g) := \inf_{\varphi \in \Lambda'} \max \left\{ \sup_{[0,1]} |\varphi - Id|, \sup_{[0,1]} d(f, g \circ \varphi) \right\}$$

for any càdlàg functions $f, g : [0, 1] \rightarrow (E, d)$. Notice that d'_S is not a distance, because it is not symmetric, but it satisfies the triangle inequality as well as the separability axiom. In other words it is a quasi-distance. Obviously we always have

$$d'_S(f, g) \leq d_S(f, g).$$

The following lemma shows that the convergence for d'_S is equivalent to the convergence for d_S when (E, d) is compact.

Lemme V.C.8

Let h and (h_n) be càdlàg functions on $[0, 1]$ with values in a metric space (E, d) . Suppose that (E, d) is compact, then

$$(d'_S(h, h_n) \xrightarrow{n \rightarrow \infty} 0 \text{ or } d'_S(h_n, h) \xrightarrow{n \rightarrow \infty} 0) \implies d_S(h_n, h) \xrightarrow{n \rightarrow \infty} 0. \tag{C.1}$$

Preuve

We show that the family (h_n) is relatively compact for the J_1 topology using an Arzelà-Ascoli-like criteria for càdlàg functions (see [49, Theorem 6.3]). For a càdlàg function x on $[0, 1]$ and $\delta > 0$ we define

$$\omega'(x, \delta) := \inf_{\{t_i\}} \max_i \sup \{d(x(s), x(t)) : s, t \in [t_{i-1}, t_i]\}.$$

where the infimum is taken over all the subdivisions $0 = t_0 < t_1 < \dots < t_n = 1$ that are δ -sparse meaning that for all i , $t_i > t_{i-1} + \delta$. By Theorem 6.3 of [49] we only need to show that


$$\limsup_{\delta \rightarrow 0} \lim_n \omega'(h_n, \delta) = 0.$$

Let $\varepsilon > 0$. By Lemma 6.2 of [49], $\lim_{\delta \rightarrow 0} \omega'(h, \delta) = 0$, thus we can find $\delta > 0$ and $\{t_i\}$ a δ -sparse partition such that for all i , and $s, t \in [t_{i-1}, t_i)$, $d(h(s), h(t)) \leq \varepsilon$. For the rest of the proof, suppose that $\lim_{n \rightarrow \infty} d_S(h, h_n) = 0$ (the case $\lim_{n \rightarrow \infty} d_S(h_n, h) = 0$ being similar). Let (φ_n) be a sequence of Λ' such that

$$\max \left\{ \sup_{[0,1]} |\varphi_n - Id|, \sup_{[0,1]} d(h, h_n \circ \varphi_n) \right\} \xrightarrow{n \rightarrow \infty} 0.$$

For n large enough, $\{\varphi_n(t_i)\}$ is δ -sparse and $\sup_{[0,1]} d(h, h_n \circ \varphi_n) \leq \varepsilon$. For $s, t \in [\varphi_n(t_{i-1}), \varphi_n(t_i))$, there exist $s', t' \in [t_{i-1}, t_i)$ such that

$$\begin{aligned} d(h_n(s), h_n(t)) &= d(h_n \circ \varphi_n(s'), h_n \circ \varphi_n(t')) \leq d(h_n \circ \varphi_n(s'), h(s')) + d(h(s'), h(t')) + d(h(t'), h_n \circ \varphi_n(t')) \\ &\leq 3\varepsilon. \end{aligned}$$

We conclude using the above mentioned criteria of relative compactity. Using (C.1) and the relative compactity, we conclude that h is an accumulation point of any sub-sequence of (h_n) for the J_1 topology, therefore (h_n) converges towards h for the J_1 topology. 

C.iv A deterministic result

In this section, we introduce precisely the excursion f_n , mentioned in Section C.i, and show the convergence of $L(f_n)$ towards $L(e)$. Write $T^n := T_1(\tau_1^n, \dots, \tau_{n-1}^n)$ the EV-labeled plane tree associated with the random decreasing factorisation $(\tau_1^n, \dots, \tau_{n-1}^n)$. Recall that for all i , $v(a_i^n)$ and $v(b_i^n)$ are the two vertices adjacent to the edge labeled i in T^n . Denote by $(S_k^n)_{0 \leq k \leq n}$ the Łukasiewicz walk associated with T^n . We complete the trajectories of S^n by linear interpolation so that S_t^n is defined for all $t \in [0, n]$. The excursion f_n is then the time renormalization of S^n , namely, $f_n(t) := S_{(n-1)t}^n$ for all $t \in [0, 1]$. Since T^n is a uniform plane tree with n vertices, S^n has the same law as a random walk, with geometric $\mathcal{G}(1/2)$ increments, conditioned on reaching -1 for the first time at n . It is then well known that f_n , when correctly renormalised in space, converges towards the Brownian excursion e . More precisely, the convergence

$$\frac{f_n}{\sqrt{2n}} \xrightarrow{n \rightarrow \infty} e \tag{C.2}$$

holds in distribution for the uniform norm. From the above convergence, one could argue that studying the process $L((2n)^{-1/2} f_n)$ instead of $L(f_n)$ would be more relevant, but one can see that $L(\alpha g) = L(g)$ for every $g \in \mathcal{E}$ and $\alpha \in \mathbb{R}_+$ so the renormalisation constant can be dropped. Notice that f_n has not necessarily unique local minima, so $L(f_n)$ has no reason to be a lamination-valued process, however it is still a càdlàg function with values in \mathbb{K} . Thus, studying its convergence in the sense of Skorokhod’s J_1 topology still makes sense.

To show that $L(f_n)$ converges in distribution towards $L(e)$ we will actually show the following deterministic convergence which, roughly speaking, shows a continuity property of the lamination-valued process with respect to the underlying excursion. Combined with (C.2) and Skorokhod’s representation theorem, it implies that $L(f_n)$ converges towards $L(e)$ in distribution in $\mathbb{D}([0, 1], \mathbb{L})$.

Proposition V.C.9

Let $g, g_n \in \mathcal{E}$ such that (g_n) converges to g for the uniform norm and g has unique local minima. Then, the convergence

$$(L_t(g_n))_{0 \leq t \leq 1} \xrightarrow{n \rightarrow \infty} (L_t(g))_{0 \leq t \leq 1}$$

holds for the Skorokhod’s J_1 topology.

Before proving Proposition V.C.9, we start with a definition and a lemma.

Définition V.C.10

- (i) For $0 \leq u \leq t \leq v \leq 1$, we denote by $[[u, t, v]] := [[u, t]] \cup [[u, v]] \cup [[t, v]]$ the *triangle* formed by the three cords with endpoints u, t and v .
- (ii) We say that $[[u, t, v]]$ is a triangle of $h \in \mathcal{E}$ if $[[u, t]]$ and $[[t, v]]$ are both cords of h (in this case $[[u, v]]$ is also a cord of h).
- (iii) For $\varepsilon > 0$ we say that the cord $[[u, v]]$ is ε -big if $v - u \geq 2\varepsilon$. Similarly, we say that the triangle $[[u, t, v]]$ is ε -big if $[[u, t]]$ and $[[t, v]]$ are both $(\varepsilon/2)$ -big.
- (iv) For an ε -big cord $[[u, v]]$ of $h \in \mathcal{E}$ we define :

$$\eta_\varepsilon(h, u, v) := \min_{[u+\varepsilon, v-\varepsilon]} h - h(u) \geq 0.$$

Lemme V.C.11

Let g and (g_n) be like in Proposition V.C.9. Fix $\varepsilon > 0$. The following assertions are true :

- (i) Let $[[u, t, v]]$ be a triangle of g such $u < t < v$. Then, for all n , there exists a triangle $[[u_n, t_n, v_n]]$ of g_n such that $\lim_{n \rightarrow \infty} (u_n, t_n, v_n) = (u, t, v)$.
- (ii) Let $\eta > 0$, then for n large enough, for every ε -big cord $[[u, v]]$ of g such that $\eta_\varepsilon(g, u, v) \geq \eta$, there exists a cord $[[u_n, v_n]]$ of g_n such that $u_n \in (u, u + \varepsilon]$ and $v_n \in [v - \varepsilon, v)$.
- (iii) There exists $\eta > 0$ such that for every ε -big cord $[[u, v]]$ of g such that $\eta_\varepsilon(g, u, v) < \eta$, there exists an ε -big cord $[[u', v']]$ of g satisfying $\eta_\varepsilon(g, u', v') = 0$, $u' \in [u, u + \varepsilon]$ and $v' \in [v - \varepsilon, v + \varepsilon]$.

Preuve du Lemme V.C.11

- (i) Let $\delta > 0$ such that u and v are not in $[t - \delta, t + \delta]$. Denote by t_n one of the points where g_n reaches its minimum over $[t - \delta, t + \delta]$. Let $u_n := \max\{a \in [0, t - \delta] : g_n(a) = g_n(t_n)\}$ and $v_n := \min\{b \in [t + \delta, 1] : g_n(b) = g_n(t_n)\}$. Then $[[u_n, t_n, v_n]]$ is a triangle of g_n . We can deduce that (u_n, t_n, v_n) converges towards (u, t, v) . Indeed, let (u', t', v') be an accumulation point of (u_n, t_n, v_n) , then by uniform convergence, $[[u', t', v']]$ is a triangle of g and t' is a minimum of g on $[t - \delta, t + \delta]$. By unicity of local minima $t' = t$. We also have that $u' \leq t - \delta \leq t + \delta \leq v'$. Once again by unicity of local minima, $u' = u$ and $v' = v$.
- (ii) Fix n such that $\|g_n - g\|_\infty < \eta/2$. Take $u_n := \max\{a \in [u, u + \varepsilon] : g_n(a) = g(u) + \eta/2\}$ and $v_n := \min\{b \in [v - \varepsilon, v] : g_n(b) = g(u) + \eta/2\}$. Then $[[u_n, v_n]]$ is a cord of g_n , $u_n \in (u, u + \varepsilon]$ and $v_n \in [v - \varepsilon, v)$.
- (iii) by contradiction, assume that for every $m \geq 1$ we can find an ε -big cord $[[u_m, v_m]]$ of g such that $\eta_\varepsilon(g, u_m, v_m) < 1/m$ and such that for every ε -big cord $[[u', v']]$ of g with $\eta_\varepsilon(g, u', v') = 0$, $u' \notin [u_m, u_m + \varepsilon]$ or $v' \notin [v_m - \varepsilon, v_m + \varepsilon]$. By compactity suppose that (u_m, v_m) converges to (u', v') . By uniform convergence $[[u', v']]$ is an ε -big cord of g such that $\eta_\varepsilon(g, u', v') = 0$. Thus for m large enough we will have $u' \in [u_m - \varepsilon, u_m + \varepsilon]$ and $v' \in [v_m - \varepsilon, v_m + \varepsilon]$. To avoid contradiction, for m large enough, $u' \in [u_m - \varepsilon, u_m)$. In other words, (v_m) converges towards v and (u_m) converges to the right towards u' without ever



touching u' . Denote by t' the moment when g reaches its minimum (which is $g(u')$) over $[u' + \varepsilon, v' - \varepsilon]$. For m large enough, $u' < u_m < t' < v_m$. Since $[[u', t']]$ and $[[u_m, v_m]]$ are both cords of g , the only possibility is that u', u_m, t' and v_m are minimum of g over $[u', v_m]$ which contradicts the uniqueness of local minima.



Preuve de la Proposition V.C.9

The idea, roughly speaking, is to perform a time change on g_n that maps all its ε -big triangles to the ε -big triangles of g , and to show that the associated lamination-valued processes are close.

The goal is to show that $\lim_{n \rightarrow \infty} d_S(L(g_n), L(g)) = 0$. For every $\varepsilon > 0$ and n large enough we will construct $\varphi_n^\varepsilon \in \Lambda$ such that, for all $\varepsilon > 0$, $\lim_n \|\varphi_n^\varepsilon - Id\|_\infty = 0$ and for all $\varepsilon > 0$ and n large enough, $\sup_{[0,1]} d_H(L_t(g), L_t(g_n \circ \varphi_n^\varepsilon)) \leq 4\pi\varepsilon$. This will show Proposition V.C.9. Indeed, notice that for all $t \in [0, 1]$, $d_H(L_t(g_n \circ \varphi_n^\varepsilon), L_{1-\varphi_n^\varepsilon(1-t)}(g_n)) \leq 2\pi\|\varphi_n^\varepsilon - Id\|_\infty$. Thus $\limsup_n d_S(L(g_n), L(g)) \leq 4\pi\varepsilon$, and since it holds for every $\varepsilon > 0$, the convergence follows.

Now, fix $\varepsilon > 0$, we tackle the construction of φ_n^ε . Notice that g has a finite number of ε -big triangles denoted by $[[u^1, t^1, v^1]], \dots, [[u^m, t^m, v^m]]$. By Lemma V.C.11 (i) we can find triangles $[[u_n^1, t_n^1, v_n^1]], \dots, [[u_n^m, t_n^m, v_n^m]]$ of g_n that converge towards the ε -big triangles of g . Let n_0 such that for all $n \geq n_0$ and for all $1 \leq i \leq m$, $u_n^i \in [0, u^i + \varepsilon]$ and such that $(u_n^1, t_n^1, v_n^1, \dots, u_n^m, t_n^m, v_n^m)$ are ordered in $[0, 1]$ like $(u^1, t^1, v^1, \dots, u^m, t^m, v^m)$. To avoid a new notation, just assume that, for all $n \geq n_0$, u_n^i is the maximum element of $[0, u^i + \varepsilon]$ such that $[[u_n^i, t_n^i]]$ is a cord of g_n . We define φ_n^ε for all $n \geq n_0$ by setting $\varphi_n^\varepsilon(u^i) = u_n^i$, $\varphi_n^\varepsilon(t^i) = t_n^i$ and $\varphi_n^\varepsilon(v^i) = v_n^i$ for all $1 \leq i \leq m$. We also set $\varphi_n^\varepsilon(0) = 0$ and $\varphi_n^\varepsilon(1) = 1$ and we complete the definition of φ_n^ε by linear interpolation. Finally, φ_n^ε is an element of Λ and clearly $\|\varphi_n^\varepsilon - Id\|_\infty$ tends to 0. For all $n \geq n_0$, define $g_n^\varepsilon := g_n \circ \varphi_n^\varepsilon$. For all $n \geq n_0$, the function g_n^ε satisfies the following property : (P1) all ε -big triangle of g is also a triangle of g_n^ε . Let $n_1 \geq n_0$ such that for all $n \geq n_1$, $\|\varphi_n^\varepsilon - Id\|_\infty \leq \varepsilon/2$. Then for all $n \geq n_1$, g_n^ε satisfies the following property : (P2) for all ε -big triangle $[[u, t, v]]$ of g there is no element $s \in (u, u + \varepsilon/2]$ such that $[[s, t]]$ is a cord of g_n^ε (this comes from the maximality of u_n^i).

It remains to show that for n large enough $\sup_{[0,1]} d_H(L_t(g), L_t(g_n \circ \varphi_n^\varepsilon)) \leq 4\pi\varepsilon$, equivalently, for n large enough the two following conditions hold : (1) for all cord $[[u, v]]$ of g there is a cord $[[u_n, v_n]]$ of g_n^ε with $u \geq u_n$ such that $[[u, v]]$ and $[[u_n, v_n]]$ are (2ε) -close and (2) for all cord $[[u_n, v_n]]$ of g_n^ε there is a cord $[[u, v]]$ of g with $u_n \geq u$ such that $[[u, v]]$ and $[[u_n, v_n]]$ are (2ε) -close. First we focus on proving (1). Let $\eta > 0$ that satisfies (iii) of Lemma V.C.11 and $n_2 \geq n_1$ such that, for all $n \geq n_2$, item (ii) of Lemma V.C.11 is satisfied with g_n replaced by g_n^ε (indeed, notice that g_n^ε converges uniformly towards g). Let $n \geq n_2$ and $[[u, v]]$ be a cord of g . There are three cases :

- If $[[u, v]]$ is not ε -big then $[[u, u]]$ is a cord of g_n^ε which is (2ε) -close to $[[u, v]]$.
- If $[[u, v]]$ is ε -big and $\eta_\varepsilon(g, u, v) \geq \eta$, then by item (ii) of Lemma V.C.11 we can find a cord $[[u_n, v_n]]$ of g_n^ε , ε -close from $[[u, v]]$, satisfying $u_n \geq u$.
- If $[[u, v]]$ is ε -big and $\eta_\varepsilon(g, u, v) < \eta$, then by item (iii) of Lemma V.C.11 we can find an ε -big cord



$[[u', v']]$ of g , ε -close from $[[u, v]]$, satisfying $u' \geq u$ and $\eta_\varepsilon(g, u', v') = 0$. Thus $[[u', v']]$ is also a cord of g_n^ε by (P1).

Now we focus on (2). By contradiction, assume that (2) fails, so there is an increasing sequence $(n_k)_k$ such that for all k , there is a cord $[[u_k, v_k]]$ of $g_{n_k}^\varepsilon$ and no cord $[[u, v]]$ of g such that $u \in [u_k, u_k + 2\varepsilon]$ and $v \in [v_k - 2\varepsilon, v_k + 2\varepsilon]$. In particular for all k , $[[u_k, v_k]]$ is ε -big since $[[u_k, u_k]]$ is a cord of g . By compactity we can suppose that (u_k, v_k) converges to (u, v) . By uniform convergence, $[[u, v]]$ is a cord of g which is ε -big. To avoid any contradiction, (u_k) converges to the right towards u . Two cases are possible : $\eta_\varepsilon(g, u, v)$ is equal to 0 or not.

- If $\eta_\varepsilon(g, u, v) > 0$, by item (ii) of Lemma V.C.11 applied with g_n replaced by g (indeed, the sequence constantly equal to g converges uniformly towards g), we can find a cord $[[u', v']]$ of g such that $u' \in (u, u + \varepsilon]$ and $v' \in [v - \varepsilon, v)$. So for k large enough $u' \in [u_k, u_k + 2\varepsilon]$ and $v' \in [v_k - 2\varepsilon, v_k + 2\varepsilon]$ which yields a contradiction.
- If $\eta_\varepsilon(g, u, v) = 0$, let $t \in [u + \varepsilon, v - \varepsilon]$ such that $[[u, t, v]]$ is an ε -big triangle of g . By (P1), $[[u, t, v]]$ is also a triangle of $g_{n_k}^\varepsilon$. For k large enough $u \leq u_k < t < v_k$. Since $[[u_k, v_k]]$, $[[u, t]]$ and $[[t, v]]$ are cords of $g_{n_k}^\varepsilon$, $[[u_k, t, v]]$ is a triangle of g_n^ε . Moreover, for k large enough, $u_k \in [u, u + \varepsilon/2]$ so by (P2), for k large enough $u_k = u$. Finally for k large enough, $u \in [u_k, u_k + 2\varepsilon]$ and $v \in [v_k - 2\varepsilon, v_k + 2\varepsilon]$ which yields a contradiction.



C.v A modification of the cord process

To show the second step of the proof of Theorem V.A.4, namely that the quasi-distance d'_ζ between L^n and $L(f_n)$ goes to 0, we introduce a new cord process $L'(f_n)$ which plays the role of an intermediate between L^n and $L(f_n)$. We start this section with the definition of $L'(f_n)$ and then show that $d'_\zeta(L'(f_n), L(f_n))$ converges towards 0 in probability using, once again, a deterministic result.

Définition V.C.12

Let $g \in \mathcal{E}$ and $[[u, v]]$ a cord of g . We say that $[[u, v]]$ is a *maximal cord* of g if there is no cord $[[u', v]]$ of g with $u' < u$ and no cord $[[u, v']]$ of g with $v' > v$. Let $\mathcal{M}(g)$ be the set of maximal cords of g , we define for all $t \in [0, 1]$,

$$L'_t(g) := [[1, 1]] \cup \overline{\bigcup_{\substack{[[u, v]] \in \mathcal{M}(g) \\ u \geq 1-t}} [[u, v]]}$$

and we write $L'(g) := (L'_t(g))_{0 \leq t \leq 1}$.

One can show that for any $g \in \mathcal{E}$ the process $L'(g)$ is càdlàg with values in \mathbb{K} . Indeed one can use the same arguments as in the proof of Proposition V.C.6 and notice the following fact : if $([[u_n, v_n]])$ is a sequence of maximal cords of g such that (u_n) converges to the right towards u and (v_n) converges towards v , then $[[u, v]]$ is either a trivial or a maximal cord of g .



Proposition V.C.13

Let $g, g_n \in \mathcal{E}$ such that (g_n) converges to g for the uniform norm. Suppose that g has unique local minima and that there is no open interval where g is monotone. Then, the following convergence holds :

$$d'_S(L'(g_n), L(g_n)) \xrightarrow{n \rightarrow \infty} 0.$$

Remarque

The Brownian excursion satisfies almost surely the hypothesis of g in Proposition V.C.13. This proposition coupled with Skorokhod's representation theorem implies the desired convergence in probability.


Preuve

Let $\varepsilon > 0$ and fix n . We will construct a function $\varphi_n^\varepsilon \in \Lambda'$ such that for n large enough, $\|\varphi_n^\varepsilon - Id\|_\infty \leq \varepsilon$ and $\sup_{t \in [0,1]} d_H(L_t(g_n), L'_{\varphi_n^\varepsilon(t)}(g_n)) \leq 4\pi\varepsilon$. For all $u \in [0,1]$, denote by $\psi_n(u)$ the minimum element of $[0, u]$ such that $[[\psi_n(u), u]]$ is a cord of g_n . Then set for all $t \in [0,1]$,

$$\varphi_n^\varepsilon(t) := 1 - (1-t) \wedge \min\{\psi_n(u) : [[u, s, v]] \text{ is an } \varepsilon\text{-big triangle of } g_n \text{ with } u \geq 1-t\}$$

with the convention that $\min \emptyset = 1$. Notice that this function is an element of Λ' . Since (g_n) converges to g and g has unique local minima, for n large enough there is no quadruplet u, t, v, w such that $[[u, t, v]]$ and $[[t, v, w]]$ are ε -big triangles of g_n (this is easily shown by contradiction). From now on suppose that n is large enough so that the previous condition is satisfied. Thus $\|\varphi_n^\varepsilon - Id\|_\infty \leq \varepsilon$. It remains to prove that for all $t \in [0,1]$, $d_H(L_t(g_n), L'_{\varphi_n^\varepsilon(t)}(g_n)) \leq 4\pi\varepsilon$. Before doing so, notice that, since g is nowhere monotone, for n large enough, there is no interval of length ε where g_n is monotone. Thus every point of $[0,1]$ is at distance at most $\varepsilon/2$ from a local maximum of g_n . Once again, from now on we suppose that n is large enough so that so previous condition is satisfied.

Let $[[u, v]]$ be a cord of g_n . We want to find a maximal cord $[[u', v']]$ of g_n which is (2ε) -close to $[[u, v]]$ such that $u' \geq 1 - \varphi_n^\varepsilon(1-u)$. If $[[u, v]]$ is not ε -big and $u \leq 1 - \varepsilon$, then there is a local maximum $u' \in [u, u + \varepsilon]$ of g_n . The trivial cord $[[u', u']]$ is maximal, (2ε) -close to $[[u, v]]$ and $u' \geq u \geq 1 - \varphi_n^\varepsilon(1-u)$. If $[[u, v]]$ is not ε -big and $u > 1 - \varepsilon$, then the cord $[[1, 1]]$ is ε -close to $[[u, v]]$. Suppose that $[[u, v]]$ is ε -big. If $\eta_\varepsilon(g_n, u, v) > 0$, then, adapting Lemma V.C.11 (ii), we can find a maximal cord $[[u', v']]$ of g_n which is ε -close to $[[u, v]]$ and such that $u < u' \leq v' < v$. Finally, suppose that $\eta_\varepsilon(g_n, u, v) = 0$. We can find a element $t \in [u + \varepsilon, v - \varepsilon]$ such that $[[u, t, v]]$ is an ε -big triangle of g_n . Taking $u' = \psi_n(u)$ and v' the maximum of $[v, 1]$ such that $[[u', v]]$ is a cord of g_n , we have that $[[u', v']]$ is a maximal cord of g_n , ε -close to $[[u, v]]$ and $u' \geq 1 - \varphi_n^\varepsilon(1-u)$.

Let $t \in [0,1]$ and $[[u', v']]$ be a maximal cord of g_n such that $u' \geq 1 - \varphi_n^\varepsilon(t)$. We want to find a cord $[[u, v]]$, (2ε) -close to $[[u', v']]$ such that $u \geq 1 - t$. Suppose that $u' < 1 - t$. There is an ε -big triangle $[[u, s, v]]$ of g_n such that $u \geq 1 - t$ and $u' \geq \psi_n(u)$. If $g_n(u') > g_n(u)$, then $v' < u$ and the trivial cord $[[u, u]]$ is ε -close to $[[u', v']]$. If $g_n(u') = g_n(u)$, then the cord $[[u, v']]$ is ε -close to $[[u', v']]$. 



C.vi Link between the cords of the decreasing factorisation and the Łukasiewicz path

In this section we finish the proof of Theorem V.A.4 by showing that $d_S(L^n, L'(f_n))$ goes to 0 in probability when n tends to infinity. Recall that $(\tau_1^n, \dots, \tau_{n-1}^n)$ denotes a uniform decreasing factorisation, that T^n denotes the associated EV-labeled plane tree, S^n its Łukasiewicz path and $f_n(t) = S_{(n-1)t}^n$.

Proposition V.C.14

The convergence

$$d_S(L^n, L'(f_n)) \xrightarrow[n \rightarrow \infty]{} 0$$

holds in probability.

Before proving this result we simplify the process $L'(f_n)$, yet another time, by removing redundant cords of f_n . More precisely, a cord $[[u, v]]$ of f_n is said to be *good* if it is a maximal cord such that v is of the form $v = i/(n-1)$ with $i \in \{0, \dots, n-1\}$. Notice that a maximal cord $[[u, v]]$ is good if and only if $f_n(u)$ is an integer. If $[[u, v]]$ is a maximal cord of f_n and $i := \lfloor v(n-1) \rfloor$, then $S_i^n - S_{i+1}^n = -1$, namely the i -th step of the Łukasiewicz path S^n is a downward step. Similarly the last step strictly before $u(n-1)$ is strictly positive. One can easily show that for any maximal cord $[[u, v]]$ of f_n there is a good cord $[[u', v']]$ of f_n , $(n-1)^{-1}$ -close to $[[u, v]]$ such that $u' \geq u$ (indeed, take $v' = \lfloor v(n-1) \rfloor / (n-1)$ and u' the smallest element such that $[[u', v']]$ is a cord of f_n).


Preuve de la Proposition V.C.14

By Skorokhod's representation theorem, suppose that (f_n) converges uniformly towards the Brownian excursion e . With this hypothesis, we will actually show that $d_S(L^n, L'(f_n))$ converges to 0 almost surely.

For all $1 \leq i \leq n-1$, denote by c_i the i -th cord of L^n , namely $c_i := [[a_i^n/n, b_i^n/n]]$. For $1 \leq i \leq n-1$, we set $s(i)$ the number of *siblings* of i , namely $s(i) := \#\{1 \leq j \leq n-1 : a_j^n = a_i^n\}$, it is the number of cords having the same left extremity than c_i . We also set $r(i)$ the *rank* of i , more precisely, $r(i) := \#\{1 \leq j \leq i : a_j^n = a_i^n\}$, it is the number of cords having the same left extremity than c_i appearing before time i . The key to the proof is to notice that for every i of rank $r(i) \geq 2$, the cord $c'_i := [[(a_i^n - h(i))/(n-1), (b_i^n - 2)/(n-1)]]$ is a good cord of f_n where $h(i) := (r(i) - 2)/(s(i) - 1)$. Moreover, all the good cords of f_n are of the form c'_i where $r(i) \geq 2$. This is a consequence of a classical property of the Łukasiewicz walk (see Figure C.1). Indeed, let $\alpha_1, \dots, \alpha_{s(i)}$ be the labels of the children of a_i^n in T^n sorted in increasing order (so in particular $\alpha_1 = a_i^n + 1$ and $\alpha_{r(i)} = b_i^n$). If we restrict the walk S^n on $[a_i^n, \alpha_{s(i)} - 1]$ then the strict descending ladder epochs correspond exactly to the times $a = \alpha_1 - 1, \dots, \alpha_{s(i)} - 1$. Since $h(i)$ is always between 0 and 1, the cord c'_i is $(2(n-1)^{-1})$ -close to c_i .

As usual we want to find a function $\varphi_n \in \Lambda$ such that $\|\varphi_n - Id\|_\infty$ and $\sup_{t \in [0,1]} d_H(L_{[nt]}^n, L'_{\varphi_n(t)}(f_n))$ tend both to 0 as n tends to infinity. What precedes gives a natural choice : for all $1 \leq i \leq n-1$ such that $r(i) \geq 2$ we set $\varphi_n(i/n) := 1 - (a_i^n - h(i))/(n-1)$. We also set $\varphi_n(0) := 0$, $\varphi_n(1) := 1$ and, as usual, we complete the definition of φ_n by linear interpolation (see Figure C.2). By Lemma V.C.15, who will follow this proof, a_i^n/n gets uniformly close to $1 - i/n$, in other words $\|\varphi_n - Id\|_\infty$ tends to 0 when n goes to infinity. Moreover, it follows from the definition that φ_n is strictly increasing.



It remains to show that for every $i \in \{1, \dots, n-1\}$, $d_H(L_i^n, L'_{\varphi^n(i/n)}(f_n))$ tends to 0 as n tends to infinity. Equivalently we need to show that every cord c_i is close to a cord c'_j with $j \leq i$ and every cord c'_j is close to a cord c_i with $i \leq j$. Most of the work has already been done, indeed for all i with $r(i) \geq 2$ the cords c_i and c'_i are $(2(n-1)^{-1})$ -close. Hence, it remains only to show that for every cord c_i with $r(i) = 1$, there is a close cord c'_j with $j \leq i$. Fix $\varepsilon > 0$, as we argued in the proof of Proposition V.C.13, for n large enough, every point in $[0, 1]$ is at distance at most ε from a local maximum of f_n . The final result readily follows, since a local maximum of f_n is a good cord (which is trivial). 

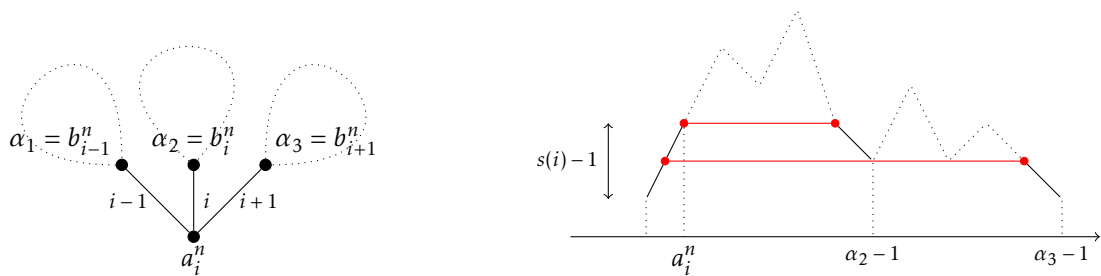


FIGURE C.1 – Illustration of Proposition V.C.14. Here $r(i) = 2$ and $s(i) = 3$. In red are represented the two good cords c'_i and c'_{i+1} of the Łukasiewicz path S^n . Those cords approximate, respectively, the cords c_i and c_{i+1} .

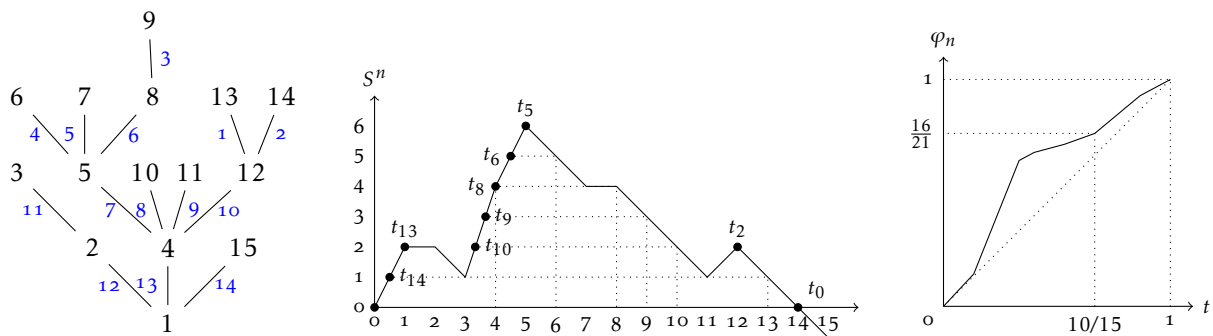


FIGURE C.2 – Construction of φ_n in an example where $n = 15$. On the left is represented the random plane tree T^{15} . The vertices are labeled in black in the lexicographic order and the edges are labeled in blue. In the middle is drawn the Łukasiewicz path S^n . The abscissa of the black dots are the values of $t_i := (1 - \varphi_n(i/n))(n-1)$ and the horizontal dotted lines represent the good cords. On the right is drawn φ_n .

The following lemma shows that $(a_{[nt]}^n/n)$ converges uniformly towards $(1-t)$ which is the last ingredient to fully prove Proposition V.C.14.

Lemme V.C.15

The convergence

$$\left(\frac{n(1-t) - a_{\lfloor nt \rfloor}^n}{\sqrt{2n}} \right)_{0 \leq t \leq 1} \xrightarrow[n \rightarrow \infty]{} \mathfrak{e} \tag{C.3}$$

holds in distribution for the uniform norm where \mathfrak{e} is the Brownian excursion on $[0, 1]$.

Preuve

We use the same notation as in the proof of Proposition V.C.14 and write $a := a_i^n$ to lighten the notations.

We show that the following relations hold

$$S_{a-1}^n \leq n - i - a = S_{a-1}^n + (s(i) - r(i)) \leq S_a^n. \tag{C.4}$$

For all $1 \leq w \leq n$ denote by k_w the number of children of the vertex labeled w in T^n (so $s(i) = k_a$). The proof of (C.4) is based on the following remark. Let $j \geq i$ then, the edge labeled j stems from a vertex smaller or equal to a for the lexicographic order. Conversely if $w < a$ then all the edges stemming from w have greater labels than i . Finally the edges stemming from a that have a greater label than i are exactly those linking a to α_s with $r(i) \leq s \leq s(i)$. Thus we have

$$n - i = \sum_{w < a} k_w + (s(i) - r(i) + 1) = S_{a-1}^n + a - 1 + (s(i) - r(i) + 1)$$

and


$$0 \leq s(i) - r(i) \leq s(i) - 1.$$

This concludes the proof of (C.4). Now we show the convergence (C.3). Let us rewrite the convergence (C.2) :

$$\left(\frac{S_{nt}^n}{\sqrt{2n}} \right)_{0 \leq t \leq 1} \xrightarrow[n \rightarrow \infty]{} \mathfrak{e}. \tag{C.5}$$

By Skorokhod's theorem, suppose that the last convergence holds almost surely. Dividing (C.4) by n we deduce that, almost surely

$$\left(\frac{a_{\lfloor nt \rfloor}^n}{n} \right)_{0 \leq t \leq 1} \xrightarrow[n \rightarrow \infty]{} (1-t)_{0 \leq t \leq 1}$$

for the uniform norm. Dividing (C.4) by $\sqrt{2n}$ then gives (C.3) (we also use the fact that the Brownian excursion is invariant in law under time inversion). 



BIBLIOGRAPHIE

- [1] ABRAHAM, R., AND DELMAS, J.-F. An introduction to Galton-Watson trees and their local limits, 2020. arXiv 1506.05571.
- [2] ADDARIO-BERRY, L., BRANDENBERGER, A., HAMDAN, J., AND KERRIOU, C. Universal height and width bounds for random trees. *Electronic Journal of Probability* 27 (2022), 1 – 24.
- [3] ADDARIO-BERRY, L., DEVROYE, L., AND JANSON, S. Sub-Gaussian tail bounds for the width and height of conditioned Galton–Watson trees. *Ann. Probab.* 41, 2 (03 2013), 1072–1087.
- [4] AIGNER, M., AND ZIEGLER, G. M. *Cayley’s formula for the number of trees*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 201–206.
- [5] ALDOUS, D. Recursive self-similarity for random trees, random triangulations and Brownian excursion. *The Annals of Probability* 22, 2 (1994), 527 – 545.
- [6] ALDOUS, D. Triangulating the circle, at random. *The American Mathematical Monthly* 101, 3 (1994), 223–233.
- [7] ALEXANDROV, A., CHAPUY, G., EYNARD, B., AND HARNAD, J. Weighted Hurwitz numbers and topological recursion : an overview. *Journal of Mathematical Physics* 59, 8 (Aug 2018), 081102.
- [8] ALON, N., AND SPENCER, J. H. *The probabilistic method*. John Wiley & Sons, 2016.
- [9] ANGEL, O., DAUVERGNE, D., HOLROYD, A. E., AND VIRÁG, B. The local limit of random sorting networks. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 55, 1 (2019), 412 – 440.
- [10] ANGEL, O., HOLROYD, A. E., ROMIK, D., AND VIRÁG, B. Random sorting networks. *Advances in Mathematics* 215, 2 (2007), 839–868.
- [11] ANGEL, O., AND MARTIN, J. 14th probability and combinatorics workshop. Private communication, Barbados 2019.
- [12] ARAUJO, F., FARINHA, J., DOMINGUES, P., SILAGHI, G., AND KONDO, D. A maximum independent set approach for collusion detection in voting pools. *Journal of Parallel and Distributed Computing* 71, 10 (2011), 1356–1366.

- [13] BANDERIER, C., KUBA, M., AND PANHOLZER, A. Analysis of three graph parameters for random trees. *Random Structures & Algorithms* 35, 1 (2009), 42–69.
- [14] BELLIN, E. Degrees in random uniform minimal factorizations. *Discret. Math.* 345 (2020), 112715.
- [15] BELLIN, E. On the independence number of random trees via tri-colourations. *33rd International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2022)*. *LIPICS* 225 (2022).
- [16] BELLIN, E. Random monotone factorisations of the cycle, 2022.
- [17] BELLIN, E. Asymptotic behaviour of the first positions of uniform parking functions. *J. Appl. Prob.* (2023). A paraître.
- [18] BERNARDI, O., AND MORALES, A. H. Counting trees using symmetries. *Journal of Combinatorial Theory, Series A* 123, 1 (2014), 104–122.
- [19] BIANE, P. Some properties of crossings and partitions. *Discrete Math.* 175 (1997), 41–53.
- [20] BIANE, P. Parking functions of types A and B. *Electron. J. Combin.* 9 (2002).
- [21] BIANE, P. Nombre de factorisations d’un grand cycle. *Séminaire Lotharingien de Combinatoire [electronic only]* 51 (2004), B51a, 4 p., electronic only–B51a, 4 p., electronic only.
- [22] BIANE, P. Nombre de factorisations d’un grand cycle. *Sem. Lothar. Combin.* 51 (2004/2005).
- [23] BIENAYMÉ, I. De la loi de multiplication et de la durée des familles. *Soc. Philomath. Paris Extraits* 5 (1845), 37–39.
- [24] BILLINGSLEY, P. *Probability and measure*. Wiley Series in Probability and Statistics, Wiley, 2012.
- [25] BOBKOV, S. G. Generalized symmetric polynomials and an approximate De Finetti representation. *Journal of Theoretical Probability* 18 (2005), 399–412.
- [26] BOLLOBÁS, B. The independence ratio of regular graphs. *Proceedings of the American Mathematical Society* 83, 2 (1981), 433–436.
- [27] BOUSQUET-MÉLOU, M., AND SCHAEFFER, G. Enumeration of planar constellations. *Advances in Applied Mathematics* 24, 4 (2000), 337–368.
- [28] BROUTIN, N., AND MARCKERT, J.-F. A new encoding of coalescent processes. Applications to the additive and multiplicative cases. *Probability Theory and Related Fields* 166, 1 (2016), 515–552.
- [29] BUTENKO, S., PARDALOS, P., SERGIENKO, I., SHYLO, V., AND STETSYUK, P. Finding maximum independent sets in graphs arising from coding theory. In *Proceedings of the 2002 ACM Symposium on Applied Computing* (New York, NY, USA, 2002), SAC ’02, Association for Computing Machinery, p. 542–546.
- [30] CARACENI, A. 14th probability and combinatorics workshop. Private communication, Barbados 2019.
- [31] CARO, Y. New results on the independence number. Tech. rep., Technical Report, Tel-Aviv University, 1979.
- [32] CAYLEY, A. A theorem on trees. *Quart. J. Pure Appl. Math.* 23 (1889), 376 – 378.
- [33] CHAPOTON, F. On some varieties associated with trees. *Michigan Mathematical Journal* 64, 4 (2015), 721 – 758.
- [34] CHASSAING, P., AND LOUCHARD, G. Phase Transition for parking blocks, Brownian excursion and coalescence. *Random Structures and Algorithms* 21 (08 2002), 76–119.



- [35] CHASSAING, P., AND MARCKERT, J.-F. Parking functions, empirical processes, and the width of rooted labeled trees. *The Electronic Journal of Combinatorics [electronic only]* 8, 1 (2001), Research paper R14, 19 p.
- [36] CHEBIKIN, D., AND POSTNIKOV, A. Generalized parking functions, descent numbers, and chain polytopes of ribbon posets. *Adv. in Appl. Math.* 44 (2010), 145–154.
- [37] CORI, R., AND ROSSIN, D. On the sandpile group of dual graphs. *European J. Combin.* 21 (2000), 447–459.
- [38] CORLESS, R., JEFFREY, D., AND KNUTH, D. A sequence of series for the Lambert W function. *ISSAC '97 : Proceedings of the 1997 international symposium on Symbolic and algebraic computation* (1997), 197–204.
- [39] COULOMB, S. Minimal vertex covers of random trees. *Journal of Statistical Mechanics : Theory and Experiment* 2005, 06 (jun 2005), P06007.
- [40] COULOMB, S., AND BAUER, M. On vertex covers, matchings and random trees, 2004.
- [41] DÉNES, J. The representation of a permutation as the product of a minimal number of transpositions, and its connection with the theory of graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* 4 (1959), 63–71.
- [42] DERSHOWITZ, N., AND ZAKS, S. Enumerations of ordered trees. *Discrete Mathematics* 31, 1 (1980), 9–28.
- [43] DEUTSCH, E. A bijection on ordered trees and its consequences. *J. Comb. Theory, Ser. A* 90 (2000), 210–215.
- [44] DEVROYE, L., GOH, M., AND ZHAO, R. The independence number of a Bienaymé-Galton-Watson tree and related parameters, 2021.
- [45] DIACONIS, P., AND HICKS, A. Probabilizing parking functions. *Advances in Applied Mathematics* 89 (2017), 125–155.
- [46] DONSKER, M. D. Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics* 23, 2 (1952), 277 – 281.
- [47] DURMIĆ, I., HAN, A., HARRIS, P. E., RIBEIRO, R., AND YIN, M. Probabilistic parking functions, 2022.
- [48] EDELMAN, P., AND GREENE, C. Balanced tableaux. *Advances in Mathematics* 63, 1 (1987), 42–99.
- [49] ETHIER, S., AND KURTZ, T. *Markov processes : characterization and convergence*. Wiley series in probability and statistics, 2005.
- [50] FÉRAY, V. Partial Jucys-Murphy elements and star factorizations. *Eur. J. Comb.* 33 (2012), 189–198.
- [51] FÉRAY, V., AND KORTCHEMSKI, I. The geometry of random minimal factorizations of a long cycle via biconditioned bitype random trees. *Ann. H. Lebesgue* 1 (2018), 149–226.
- [52] FÉRAY, V., AND KORTCHEMSKI, I. Trajectories in random minimal transposition factorizations. *Lat. Am. J. Probab. Math. Stat.* 16 (2019), 759–785.
- [53] FÉRAY, V., LOUF, B., AND THÉVENIN, P. Random generation and scaling limits of fixed genus factorizations into transpositions, 2021. Preprint available on arXiv, arXiv :2105.03284.
- [54] FUCHS, M., HOLMGREN, C., MITSCHKE, D., AND NEININGER, R. A note on the independence number, domination number and related parameters of random binary search trees and random recursive trees. *Discrete Applied Mathematics* 292 (2021), 64–71.
- [55] GALTON, F., AND WATSON, H. On the probability of extinction of families. *J. Anthropol. Inst.* 4 (1874), 138—144.



- [56] GAREY, M., AND JOHNSON, D. *Computers and intractability*, vol. 174. freeman San Francisco, 1979.
- [57] GEWURZ, D. A., AND MEROLA, F. Some factorisations counted by Catalan numbers. *Eur. J. Comb.* 27 (2006), 990–994.
- [58] GOULDEN, I., AND YONG, A. Tree-like properties of cycle factorizations. *J. Combinatorial Theory ser. A* 98 (2002), 106–117.
- [59] GOULDEN, I. P., AND JACKSON, D. M. Transitive powers of Young–Jucys–Murphy elements are central. *Journal of Algebra* 321, 7 (2009), 1826–1835.
- [60] GOULDEN, I. P., AND PEPPER, S. Labelled trees and factorizations of a cycle into transpositions. *Discrete Math.* 113 (1993), 263–268.
- [61] HURWITZ, A. Uber Riemann’sche flächen mit gegebenen verzweigungspunkten. *Math. Ann.* 39 (1891), 1–66.
- [62] IBRAGIMOV, I. A., AND LINNIK, Y. V. *Independent and stationary sequences of random variables*. Wolters-Noordhoff, Groningen, 1971.
- [63] IRVING, J., AND RATTAN, A. Minimal factorizations of permutations into star transpositions. *Discret. Math.* 309, 6 (2009), 1435–1442.
- [64] IRVING, J., AND RATTAN, A. Trees, parking functions and factorizations of full cycles. *European Journal of Combinatorics* 93 (03 2021), 103257.
- [65] ITO, K., AND McKEAN, H. P. Diffusion processes and their sample paths. *Journal of the American Statistical Association* 91 (1996), 1754.
- [66] JACOD, J., AND SHIRYAEV, A. N. *Limit theorems for stochastic processes*, vol. 288 of Grundlehren der Mathematischen Wissenschaften. Springer-Verlag (second edition), 2003.
- [67] JANSON, S. Random cutting and records in deterministic and random trees. *Random Structures Algorithms* 29, 2 (2006), 139–179.
- [68] JANSON, S. Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation. *Probability Surveys* 9, none (2012), 103 – 252.
- [69] JANSON, S. Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees. *Random Struct. Algorithms* 48 (2016), 57–101.
- [70] JANSON, S. On the independence number of some random trees. *Electronic Communications in Probability* 25, none (2020), 1 – 14.
- [71] JOO, C., LIN, X., J.RYU, AND SHROFF, N. Distributed greedy approximation to maximum weighted independent set for scheduling with fading channels. In *Proceedings of the fourteenth ACM International Symposium on Mobile ad hoc Networking and Computing* (New York, NY, USA, 2013), MobiHoc ’13, Association for Computing Machinery, p. 89–98.
- [72] KARP, R. M. *Reducibility among combinatorial problems*. Springer US, Boston, MA, 1972, pp. 85–103.
- [73] KENYON, R., AND YIN, M. Parking functions : from combinatorics to probability, 2021. arXiv 2103.17180.
- [74] KESTEN, H. Subdiffusive behavior of random walk on a random cluster. *Annales de l’I.H.P. Probabilités et statistiques* 22, 4 (1986), 425–487.



- [75] KONHEIM, A. G., AND WEISS, B. An occupancy discipline and applications. *SIAM J. Appl. Math.* 14 (1966), 1266–1274.
- [76] KORTCHEMSKI, I. Limit theorems for conditioned non-generic Galton–Watson trees. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 51, 2 (2015), 489 – 511.
- [77] KÖNIG, D. Gráfok és Mátrixok. *Matematikai és Fizikai Lapok* (1931), 116–119.
- [78] LE GALL, J.-F. Random trees and applications. *Probability Surveys* 2, none (2005), 245–311.
- [79] LOKSHTANOV, D., VATSHELLE, M., AND VILLANGER, Y. Independent set in p_5 -free graphs in polynomial time. *Proceedings of SODA* (01 2014), 570–581.
- [80] LUBY, M. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing* 15, 4 (1986), 1036–1053.
- [81] LUPU, T., PITMAN, J., AND TANG, W. The Vervaat transform of Brownian bridges and Brownian motion. *Electronic Journal of Probability* 20 (10 2013).
- [82] MATSUMOTO, S., AND NOVAK, J. Unitary matrix integrals, primitive factorizations, and Jucys-Murphy Elements. *22nd International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2010). DMTCS Proceedings* (2010), 403–412.
- [83] MEIR, A., AND MOON, J. W. On the altitude of nodes in random trees. *Canadian Journal of Mathematics* 30 (1978), 997–1015.
- [84] MINTY, G. On maximal independent sets of vertices in claw-free graphs. *Journal of Combinatorial Theory, Series B* 28, 3 (1980), 284–304.
- [85] MOSZKOWSKI, P. A solution to a problem of Dénes : a bijection between trees and factorizations of cyclic permutations. *European J. Combin.* 10 (1989), 13–16.
- [86] NEVEU, J. Arbres et processus de Galton-Watson. In *Annales de l’IHP probabilités et statistiques* (1986), vol. 22, pp. 199–207.
- [87] PAK, I. Reduced decompositions of permutations in terms of star transpositions, generalized Catalan numbers and k -ARY trees. *Discret. Math.* 204, 1-3 (1999), 329–335.
- [88] PITMAN, J. Combinatorial stochastic processes. *Lecture Notes in Mathematics* 1875 (2006). Springer, New York.
- [89] PRUFER, H. Neuer Beweis eines Satzes über Permutationen. *Arch. Math. Phys.* 27 (1918), 742–744.
- [90] RIORDAN, J. *Combinatorial Identities*. New York : Wiley, 1979.
- [91] ROBSON, J. Algorithms for maximum independent sets. *Journal of Algorithms* 7, 3 (1986), 425–440.
- [92] SHI, J. Y. The Kazhdan-Lusztig cells in certain affine Weyl groups. *Lecture Notes in Mathematics* 1179 (1986). Springer-Verlag, Berlin.
- [93] STANLEY, R. P. On the number of reduced decompositions of elements of Coxeter groups. *European Journal of Combinatorics* 5, 4 (1984), 359–372.
- [94] STANLEY, R. P. Parking functions and noncrossing partitions. *Electron. J. Combin.* 4, Research Paper 20 (1997). 20 pp.
- [95] STANLEY, R. P. Hyperplane arrangements, parking functions and tree inversions. *Mathematical Essays in Honor of Gian-Carlo Rota. Progr. Math.* 161 (1998), 359–375. Birkhauser, Boston.



-
- [96] STANLEY, R. P., AND PITMAN, J. A polytope related to empirical distributions, plane trees, parking functions, and the associahedron. *Discrete Comput. Geom* 27 (2002), 603–634.
- [97] STUFLER, B. Local limits of large Galton–Watson trees rerooted at a random vertex. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 55, 1 (2019), 155 – 183.
- [98] THÉVENIN, P. A geometric representation of fragmentation processes on stable trees. *The Annals of Probability* 49, 5 (2021), 2416–2476.
- [99] WEI, V. K. A lower bound on the stability number of a simple graph. Tech. rep., Bell Laboratories Technical Memorandum New Jersey, 1981.
- [100] WELLNER, J. A. The Cramér-Chernoff method and some exponential bounds.
- [101] XIAO, M., AND NAGAMACHI, H. Confining sets and avoiding bottleneck cases : a simple maximum independent set algorithm in degree-3 graphs. *Theor. Comput. Sci.* 469 (2013), 92–104.
- [102] XIAO, M., AND NAGAMACHI, H. Exact algorithms for maximum independent set. *Information and Computation* 255 (2017), 126–146.
- [103] YAN, C. H. Parking functions. *Handbook of Enumerative Combinatorics. Discrete Math. Appl.*, CRC Press, Boca Raton, FL 2015 (1986), 835–893.
- [104] YIN, M. Parking functions : interdisciplinary connections, 2021. arXiv 2107.01767.
- [105] ZITO, J. The structure and maximum number of maximum independent sets in trees. *Journal of Graph Theory* 15, 2 (1991), 207–221.

Titre : Structures arborescentes aléatoires

Mots clés : Arbres et marches aléatoires, arbres de Bienaymé-Galton-Watson, factorisations minimales et monotones du cycle, processus de lamination, tuples de stationnement

Résumé : Cette thèse se concentre sur certaines structures arborescentes aléatoires et leurs applications. Quatre modèles sont étudiés. Le premier est celui des factorisations minimales du cycle de taille n . Une telle factorisation est la donnée de $n - 1$ transpositions dont la composition égale la permutation cyclique $(1 \dots n)$. En exploitant une bijection avec les arbres étiquetés et en explicitant certaines fonctions génératrices sur ces arbres, nous déduisons la loi jointe du nombre de fois que 1 et 2 apparaissent dans une factorisation minimale de taille n choisie uniformément au hasard. Le deuxième problème est celui des tuples de stationnement (*parking functions*). Nous étudions la distance entre les k premières coordonnées d'un n -uplet de stationnement choisi uniformément au hasard et k variables i.i.d uniformes. Nous montrons que la distance en variation totale tend vers 0 lorsque $k = o(\sqrt{n})$ et que la distance de Kolmogorov tend vers 0 lorsque $k = o(n)$. Nous donnons aussi des estimations sur la vitesse de convergence. Cette étude est possible, encore une fois, grâce à une bijection avec les arbres étiquetés qui permet

de se ramener à l'étude d'arbres aléatoires puis de marches aléatoires conditionnées. Le troisième problème consiste à étudier le nombre d'indépendance d'arbres aléatoires. Le nombre d'indépendance d'un graphe est le plus grand cardinal d'un sous-ensemble de sommets du graphe tel qu'aucune paire de sommets de cet ensemble ne sont reliés par une arête. Nous obtenons des théorèmes limites pour les arbres simplement générés par des méthodes robustes de convergence locale, et généralisons des résultats précédemment obtenus de manière calculatoire. Enfin le dernier problème est celui des factorisations monotones qui sont des factorisations minimales satisfaisant une certaine condition supplémentaire de monotonie. Nous décrivons une bijection entre ces objets et les arbres plans. Pour tout $k < n$ nous représentons les k premières transpositions d'une factorisation monotone de taille n comme une lamination du disque, i.e comme une union de cordes disjointes du cercle unité. Nous étudions ensuite la convergence du processus $(0 \leq k < n)$ de lamination en exploitant la bijection avec les arbres plans.

Title : Random tree structures

Keywords : Random walks and trees, Bienaymé-Galton-Watson trees, minimal and monotone factorisations of the cycle, lamination process, parking functions

Abstract : This thesis focuses on some random tree structures and their applications. Four models are studied. The first one is about minimal factorisations of the cycle of size n . Such a factorisation is given by $n - 1$ transpositions whose composition gives the cyclic permutation $(1 \dots n)$. Taking advantage of a bijection with labelled trees and computing some generating functions, we deduce the joint law of the number of times 1 and 2 appear in a minimal factorisation of size n chosen uniformly at random. The second problem is about parking functions. We study the distance between the first k coordinates of a n -parking function chosen uniformly at random and k i.i.d uniform random variables. We show that the total variation distance tends to 0 when $k = o(\sqrt{n})$ and the Kolmogorov distance goes to 0 when $k = o(n)$. We also provide some estimate for the speed of convergence. Those results are possible, once again, thanks to a

bijection with labelled trees which allows to transpose the study towards random trees and random walks. The third problem is about the Independence number of random trees. The independence number of a graph is the biggest cardinal of a subset of vertices such that no pair of vertices are adjacent. We obtain limit theorems for simply generated trees using local convergence methods and generalise some previous results. Lastly, the fourth problem is about monotone factorisations which are minimal factorisation satisfying an extra condition of monotonicity. We describe a bijection between those objects and plane trees. For all $k < n$, we represent the first k transpositions of a monotone factorisation of size n as a lamination of the disk, i.e as a union of non-crossing cords. Then, we study the convergence of the lamination process $(0 \leq k < n)$ exploiting the previous bijection with plane trees.