
FONCTIONNELLES ADDITIVES SUR LES ARBRES ALÉATOIRES

1.1 Arbres discrets

Un arbre discret est un graphe sans cycle, non orienté et connexe (on peut toujours relier deux nœuds de l'arbre par un chemin). Dans la suite, nous allons considérer des arbres enracinés et ordonnés. Un arbre est dit enraciné si un nœud est distingué comme étant la racine notée \emptyset . Il est alors possible de décrire un arbre par sa suite des générations : la racine est la génération 0, les voisins de la racine forment la génération 1 et plus généralement, les nœuds à distance k de la racine forment la k -ième génération. Si un nœud u de la génération k a des voisins dans la génération $k + 1$ alors ces voisins sont appelés les enfants de u . Un arbre est dit ordonné si les enfants de chaque nœud sont munis d'un ordre. Autrement dit, les enfants du nœud u peuvent être ordonnés dans une suite u_1, \dots, u_d où d est le nombre d'enfants de u . Les arbres enracinés et ordonnés sont aussi parfois appelés arbres planaires.

1.1.1 Notions sur les arbres discrets

Nous allons définir les arbres planaires par leur ensemble de nœuds en utilisant le formalisme de Neveu [153]. On désigne par $\mathcal{U} = \bigcup_{n \geq 0} (\mathbb{N}^*)^n$ l'ensemble des suites finies d'entiers strictement positifs. Par convention, $(\mathbb{N}^*)^0 = \{\emptyset\}$. Pour $n \geq 0$ et $u \in (\mathbb{N}^*)^n \subset \mathcal{U}$, la longueur de u est notée $|u| = n$. Si $u = (u_1, u_2, \dots, u_n) \in \mathcal{U} \setminus \{\emptyset\}$, alors on appelle $\tilde{u} = (u_1, u_2, \dots, u_{n-1})$ le parent de u (u est alors un enfant de \tilde{u}). Soient $v, w \in \mathcal{U}$. La concaténation des suites v et w est notée vw avec la convention que $v\emptyset = v$ et $\emptyset w = w$. On dit que v est un ancêtre de u (au sens large) et on écrit $v \preceq u$ s'il existe $w \in \mathcal{U}$ tel que $u = vw$. Si $v \preceq u$ et $v \neq u$, alors on écrit $v \prec u$. Ainsi, \preceq désigne l'ordre généalogique sur \mathcal{U} . On peut également définir l'ordre lexicographique sur \mathcal{U} noté $<$. Par exemple, $\emptyset < 1 < (1, 2) < (2, 1) < (2, 2)$. Remarquons que $u \prec v$ implique que $u < v$ mais la réciproque est fautive ($(1, 2) < (1, 3)$ mais $(1, 2) \not\prec (1, 3)$).

Définition. *Un arbre planaire \mathbf{t} est un sous-ensemble de \mathcal{U} satisfaisant les trois propriétés suivantes :*

- *Un arbre contient toujours la racine : $\emptyset \in \mathbf{t}$,*
- *Si un arbre contient un individu $u \in \mathbf{t} \setminus \{\emptyset\}$ alors il contient son parent \tilde{u} : si $u \in \mathbf{t} \setminus \{\emptyset\}$, alors $\tilde{u} \in \mathbf{t}$.*
- *Chaque individu a un nombre fini d'enfants : pour tout $u \in \mathbf{t}$, il existe $k_u(\mathbf{t}) \in \mathbb{N}$ tel que, pour chaque $i \in \mathbb{N}^*$, $ui \in \mathbf{t}$ si et seulement si $1 \leq i \leq k_u(\mathbf{t})$.*

Pour $u \in \mathbf{t}$, l'entier $k_u(\mathbf{t})$ représente le nombre de descendants du nœud u . Le nœud u est appelée feuille (resp. nœud interne) si $k_u(\mathbf{t}) = 0$ (resp. $k_u(\mathbf{t}) > 0$). Le nœud \emptyset est appelé la racine de \mathbf{t} . On note $|\mathbf{t}| = \text{Card}(\mathbf{t})$ le nombre de nœuds de \mathbf{t} et on dit que \mathbf{t} est fini si $|\mathbf{t}| < +\infty$. On note \mathbb{T} l'ensemble des arbres planaires, \mathbb{T}_0 l'ensemble des arbres planaires finis et $\mathbb{T}^{(p)} = \{\mathbf{t} \in \mathbb{T}, |\mathbf{t}| = p\}$ l'ensemble des arbres planaires à p nœuds. Rappelons que le nombre d'arbres planaires à p nœuds est donné par le $p - 1$ -ième nombre de Catalan (voir Drmota [68], partie 1.2.2) :

$$C_{p-1} = \frac{1}{p} \binom{2p-2}{p-1} = \frac{(2p-2)!}{(p-1)!p!}.$$

La hauteur H d'un arbre $\mathbf{t} \in \mathbb{T}$ est définie par :

$$H(\mathbf{t}) = \sup\{|u| : u \in \mathbf{t}\}.$$

On définit également le sous-arbre $\mathbf{t}_u \in \mathbb{T}$ de \mathbf{t} « au dessus » de u par :

$$\mathbf{t}_u = \{v \in \mathbf{t}, uv \in \mathbf{t}\}.$$

Ces sous-arbres sont appelés « fringe-trees » en anglais, ils ont été introduits par Aldous [7] (voir aussi Holmgren et Janson [109]). La figure 1.1 donne un exemple de sous-arbre.

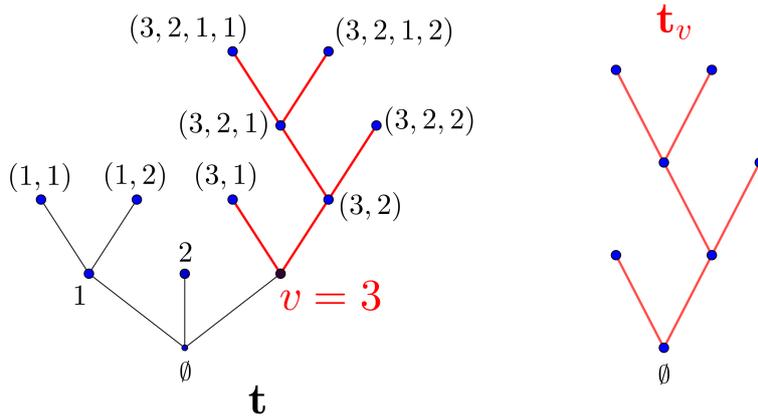


FIGURE 1.1 – Un arbre planaire à 5 nœuds internes et son sous-arbre issu du nœud $v = 3$ dessiné en rouge à droite de la figure.

On peut coder un arbre planaire fini par des fonctions, voir figure 1.2 (voir par exemple Le Gall [135]). Commençons par la fonction de contour (ou chemin de Dick). Pour cela, rappelons en quoi consiste le parcours d'un arbre en profondeur. Il s'agit de parcourir l'arbre, en partant de la racine \emptyset au temps $t = 0$, et, arrivé en un nœud u , d'aller vers le premier enfant non visité u_i ou alors, si tous les enfants ont déjà été visités, de retourner vers son parent. Le procédé s'arrête lorsqu'on est revenu à la racine en ayant visité tous les enfants de celle-ci. Étant donné que chaque arête est visitée deux fois, il est clair que le temps total d'exploration de l'arbre est $2(|\mathbf{t}| - 1)$. Pour $\mathbf{t} \in \mathbb{T}$, la fonction de contour de \mathbf{t} est la fonction $C^{\mathbf{t}} : [0, 2(|\mathbf{t}| - 1)] \rightarrow \mathbb{R}^+$ qui associe à chaque étape $k \in \llbracket 0, 2(|\mathbf{t}| - 1) \rrbracket$ du parcours en profondeur, la génération de l'individu v_k visité. On interpole ensuite linéairement $C^{\mathbf{t}}$ sur l'intervalle $[0, 2(|\mathbf{t}| - 1)]$, et l'on pose $C^{\mathbf{t}}(t) = 0$ pour tout $t > 2(|\mathbf{t}| - 1)$.

On peut également coder un arbre par sa fonction de hauteur. La fonction de hauteur de $\mathbf{t} \in \mathbb{T}$ est la suite de générations des individus de \mathbf{t} , quand ces individus sont classés dans l'ordre lexicographique. Si $\mathbf{t} \in \mathbb{T}$, on énumère les sommets dans l'ordre lexicographique $u_0 = \emptyset < u_1 < \dots < u_{|\mathbf{t}|-1}$. Pour chaque $n \in \llbracket 0, |\mathbf{t}| - 1 \rrbracket$, on définit $h_{\mathbf{t}}(n) = |u_n|$ comme la hauteur du sommet u_n . On pose $h_{\mathbf{t}}(m) = 0$ pour $m \geq |\mathbf{t}|$, et on prolonge alors $h_{\mathbf{t}}$ à \mathbb{R}^+ par

interpolation linéaire. La fonction $h_{\mathbf{t}} = (h_{\mathbf{t}}(t), t \geq 0)$ est appelée fonction de hauteur de \mathbf{t} . La fonction de hauteur est similaire à la fonction de contour mais chaque individu de l'arbre n'est visité qu'une fois.

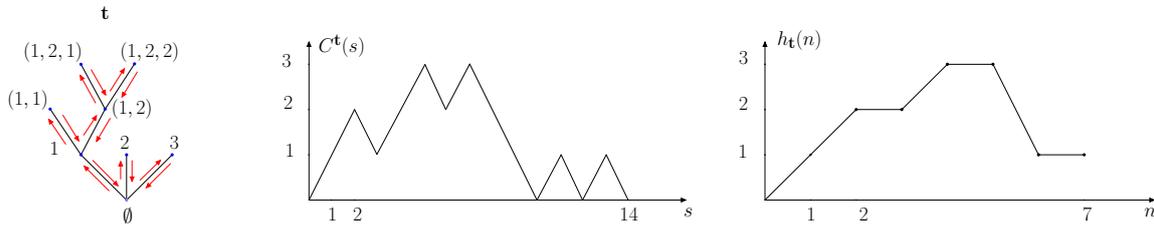


FIGURE 1.2 – Un arbre \mathbf{t} , son processus de contour $C^{\mathbf{t}}$ et sa fonction de hauteur $h_{\mathbf{t}}$

Dans la suite nous allons nous intéresser à deux types d'arbres aléatoires : les arbres aléatoires binaires (auxquels on associe généralement deux modèles classiques, le modèle de Catalan et le modèle de permutations aléatoires) et les arbres simplement générés (et plus particulièrement les arbres de Galton-Watson conditionnés par leur taille totale).

Nous commençons par quelques rappels sur les arbres binaires. Le lecteur pourra se référer au livre de Drmota [68], parties 1.2.1 et 1.4.1, Mahmoud [142], chapitre 2 et Sedgewick et Flajolet [177], chapitre 6. Ces arbres ont de nombreuses applications notamment en informatique au travers des algorithmes de tri. Ces algorithmes ont été étudiés à la fin des années 1960 et au début des années 1970 par Knuth [126, 128], qui donne des analyses détaillées de plusieurs paramètres combinatoires qui permettent de déterminer par exemple leur performance moyenne. Sedgewick [176], quant à lui, s'est intéressé à un des modèles de tri les plus utilisés, l'algorithme Quicksort, qui a été inventé par Hoare au début des années 1960.

1.1.2 Arbres binaires

Définition

Un arbre binaire complet est un arbre enraciné et ordonné où chaque nœud a zéro (c'est une feuille) ou deux enfants (c'est un nœud interne). Un arbre binaire à n nœuds internes a $n + 1$ nœuds externes et donc $2n + 1$ nœuds au total. On note $\mathbb{T}_{\text{bin}}^{(n)}$ l'ensemble des arbres binaires complets enracinés à n nœuds internes.

Arbres binaires sous le modèle de Catalan

Le plus simple pour choisir au hasard un arbre dans un ensemble d'arbres est le choix uniforme. Ce modèle est aussi appelé modèle de Catalan. Il sera l'objet de notre étude dans le chapitre 3. Soit T_n une variable aléatoire à valeurs dans $\mathbb{T}_{\text{bin}}^{(n)}$ choisie uniformément dans l'ensemble fini $\mathbb{T}_{\text{bin}}^{(n)}$. Comme le nombre d'arbres binaires complets avec n nœuds internes est donné par le n -ième nombre de Catalan $C_n = \frac{1}{n+1} \binom{2n}{n}$, on obtient que la probabilité d'obtenir un arbre particulier est l'inverse du n -ième nombre de Catalan. La loi de T_n est donnée par

$$\mathbb{P}(T_n = \mathbf{t}) = \frac{1}{C_n}, \text{ pour tout } \mathbf{t} \in \mathbb{T}_{\text{bin}}^{(n)}.$$

Le choix uniforme d'un arbre parmi un ensemble d'arbres peut également s'appliquer à d'autres types d'arbres que les arbres binaires. Par analogie, on parlera encore de « modèle de Catalan » à ne pas confondre avec ce qu'on appelle les arbres de Catalan qui sont des arbres planaires.

Arbres marqués

Dans la suite, il conviendra de faire la distinction entre deux types d'arbres : les arbres non marqués et les arbres marqués. Dans le premier cas, les nœuds des arbres ne contiennent pas d'information et l'on s'intéresse donc simplement à leur forme. Dans le second cas, les arbres sont vus comme des structures de données c'est-à-dire que les nœuds contiennent des informations aussi appelées clés ou marques. Si l'on efface les clés d'un arbre marqué, l'arbre non marqué obtenu est appelé sa forme. Formellement, un arbre marqué est une paire $\tilde{\mathbf{t}} = (\mathbf{t}, (\gamma_v, v \in \mathbf{t}))$ où les marques γ_v sont des éléments d'un ensemble Γ . L'arbre non marqué \mathbf{t} est appelé le squelette ou la forme de l'arbre marqué $\tilde{\mathbf{t}}$, voir la définition donnée par Neveu [153]. C'est l'ensemble des nœuds de l'arbre marqué.

Arbres binaires de recherche et modèle de permutations uniformes

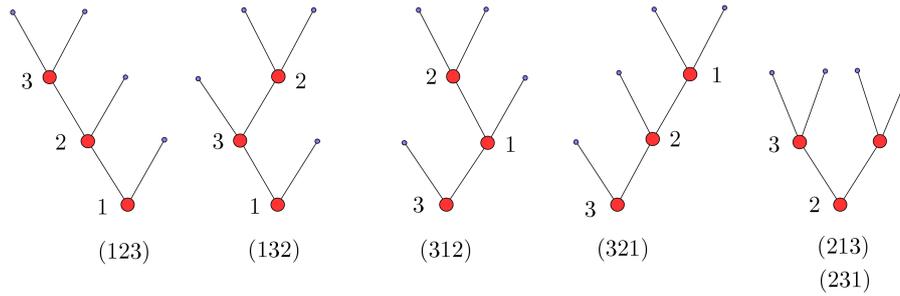
Une des applications les plus importantes des arbres binaires est l'algorithme des arbres binaires de recherche (ABR en abrégé ou encore BST pour « binary search tree » en anglais), voir Mahmoud [142], chapitre 2 ou Chauvin, Clément et Gardy [47], partie 1.2.5. Les ABR sont des structures de données fondamentales en informatique. Ils sont utiles pour classer et chercher des données, appelées des clés. Un ABR à n nœuds internes est un arbre binaire marqué par des clés x_1, \dots, x_n prises dans un ensemble totalement ordonné Γ qui satisfait la contrainte que la clé de chaque nœud interne est plus grande que toutes les clés de son sous-arbre de gauche et plus petite que toutes les clés de son sous-arbre de droite. On peut rajouter $n + 1$ feuilles non étiquetées à cette structure afin d'obtenir un arbre binaire complet marqué à n nœuds internes. Le squelette d'un ABR à n nœuds internes est donc un arbre binaire complet à n nœuds internes et donc $n + 1$ feuilles et $2n + 1$ nœuds.

Une généralisation de ces arbres sont les arbres m -aire de recherche (en anglais, m -ary search trees), voir Mahmoud [142], partie 1.7, Drmota [68], partie 1.4.2 ou Holmgren et Janson [109], où chaque nœud peut contenir jusqu'à $m - 1$ éléments d'un ensemble totalement ordonné.

On peut munir l'ensemble des ABR à n nœuds internes d'une loi de probabilité. Le modèle le plus étudié dans la littérature est le modèle de permutations uniformes (RPM en anglais pour « random permutation model »), voir Fill [85], Mahmoud [142], chapitre 2 ou Chauvin, Clément et Gardy [47], partie 2.2.1. Un ABR à n nœuds internes peut être généré par une permutation (ou liste) π de l'ensemble $\llbracket 1, n \rrbracket$. A chaque permutation de $\llbracket 1, n \rrbracket$ est associé un ABR. Dans le modèle de permutations uniformes, chaque permutation π de l'ensemble $\llbracket 1, n \rrbracket$ est équiprobable. Remarquons que des permutations différentes peuvent engendrer le même arbre squelette (voir figure 1.3). Ainsi, contrairement au modèle de Catalan, tous les arbres squelettes ne sont pas équiprobables. Soit $\tilde{\mathbf{T}}_n$ une variable aléatoire à valeurs dans l'ensemble des arbres binaires complets marqués choisie selon le modèle de permutations uniformes. On note \mathbf{T}_n l'arbre squelette associé à $\tilde{\mathbf{T}}_n$ qui est une variable aléatoire à valeurs dans $\mathbb{T}_{\text{bin}}^{(n)}$. La loi de \mathbf{T}_n est donnée par

$$Q(\mathbf{t}) = \mathbb{P}(\mathbf{T}_n = \mathbf{t}) = \frac{1}{\prod_{v \in \mathbf{t}} |\mathbf{t}_v|}, \text{ pour tout } \mathbf{t} \in \mathbb{T}_{\text{bin}}^{(n)}. \quad (1.1)$$

L'ABR aléatoire issu d'une permutation uniforme de taille n a une forme très différente d'un arbre binaire de Catalan : sa hauteur (resp. sa longueur totale, ou longueur de cheminement, égale à la somme des distances des différents nœuds à la racine) est asymptotiquement d'ordre $\log n$ (resp. d'ordre $n \log n$) alors qu'elle est asymptotiquement d'ordre \sqrt{n} (resp. d'ordre $n\sqrt{n}$) sous le modèle de Catalan. Ainsi, le modèle de permutations uniformes donne plus de poids aux arbres dits équilibrés que le modèle de Catalan (voir figure 1.3).

FIGURE 1.3 – $C_3 = 5$ arbres binaires complets ordonnés et enracinés à 3 nœuds internes

Faisons maintenant quelques rappels concernant les arbres de Galton-Watson qui constituent une classe importante d'exemples d'arbres aléatoires ordonnés et enracinés.

1.2 Arbres de Galton-Watson

1.2.1 Définitions

Soit $\mathbf{p} = (\mathbf{p}(n), n \in \mathbb{N})$ une loi de probabilité sur \mathbb{N} . Le processus de Galton-Watson $Z = (Z_n, n \in \mathbb{N})$ de loi de reproduction \mathbf{p} décrit l'évolution de la taille d'une population issue d'un individu et dont les individus ont chacun un nombre d'enfants suivant la loi \mathbf{p} , de façon indépendante les uns des autres. Autrement dit, si on introduit une famille $(\xi_{i,n}; i \in \mathbb{N}, n \in \mathbb{N})$ de variables aléatoires indépendantes de loi \mathbf{p} , on définit le processus Z par récurrence par

$$Z_0 = 1 \quad \text{et} \quad Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n+1}.$$

L'arbre de Galton-Watson associé au processus de Galton-Watson Z représente l'arbre généalogique de la population où chaque individu est relié à son parent et à ses descendants. Z_n représente le nombre de nœuds au niveau k dans l'arbre de Galton-Watson et $\sum_{k \geq 0} Z_k$ le nombre de nœuds de l'arbre de Galton-Watson :

Définition. Soit \mathbf{p} une mesure de probabilité sur \mathbb{N} . Soit τ une variable aléatoire à valeurs dans \mathbb{T} . On dit que τ est un arbre de Galton-Watson de loi de reproduction \mathbf{p} si l'on a :

- $\mathbb{P}(k_\emptyset(\tau) = j) = \mathbf{p}(j)$ pour tout $j \in \mathbb{N}$,
- pour tout $j \in \mathbb{N}^*$, sous $\mathbb{P}(\tau | k_\emptyset(\tau) = j)$, les j sous-arbres τ_1, \dots, τ_j de τ issus de \emptyset sont indépendants et de même loi que τ .

On peut montrer que cette loi existe et est uniquement caractérisée par \mathbf{p} . En fait, τ est un arbre de Galton-Watson de loi de reproduction \mathbf{p} si et seulement si pour tout $h \in \mathbb{N}^*$ et $\mathbf{t} \in \mathbb{T}$ tel que $H(\mathbf{t}) \leq h$, on a :

$$\mathbb{P}(r_h(\tau) = \mathbf{t}) = \prod_{\substack{u \in \mathbf{t} \\ |u| < h}} \mathbf{p}(k_u(\mathbf{t})),$$

où pour tout $\mathbf{t} \in \mathbb{T}$ et pour tout $h \in \mathbb{N}^*$, $r_h(\mathbf{t})$ est le sous-arbre de \mathbf{t} obtenu en gardant tous les nœuds à hauteur inférieure ou égale à h et est défini par : $r_h(\mathbf{t}) = \{u \in \mathbf{t} : |u| \leq h\}$. En particulier, quand $\tau \in \mathbb{T}_0$, on a la formule suivante :

$$\mathbb{P}(\tau = \mathbf{t}) = \prod_{u \in \mathbf{t}} \mathbf{p}(k_u(\mathbf{t})), \quad \text{pour tout } \mathbf{t} \in \mathbb{T}_0. \quad (1.2)$$

La probabilité $\mathbb{P}(\tau = \mathbf{t})$ est ce qu'on appellera le poids de \mathbf{t} dans le cadre des arbres simplement générés.

On exclut le cas trivial où $\mathbf{p}(1) = 1$. On rappelle que le processus de Galton-Watson Z est dit sous-critique (resp. critique, sur-critique) si le nombre moyen d'enfants $m = \sum_{k \in \mathbb{N}} k\mathbf{p}(k)$ satisfait $m < 1$ (resp. $m = 1$, $m > 1$). Un arbre de Galton-Watson de loi de reproduction \mathbf{p} est fini presque-sûrement si $m \leq 1$ et infini avec probabilité strictement positive si $m > 1$ (voir Harris [102], chapitre 1 ou Athreya et Ney [15], chapitre 3).

1.2.2 Arbres de Galton-Watson conditionnés et arbres simplement générés

Dans cette partie, nous allons nous intéresser aux arbres de Galton-Watson conditionnés à avoir un certain nombre de nœuds. Nos résultats asymptotiques sur les fonctionnelles additives du Chapitre 3 concerneront en particulier ce modèle. En fait, les arbres de Galton-Watson conditionnés rentrent dans le cadre plus général des arbres simplement générés. Nous précisons les liens qui existent entre ces deux modèles d'arbres. Nous renvoyons à Drmota [68], partie 1.2.7 et Janson [114], chapitre 2 pour les définitions et des propriétés des arbres de Galton-Watson conditionnés et des arbres simplement générés.

Les arbres de Galton-Watson ont des tailles aléatoires. On peut cependant s'intéresser à des arbres aléatoires avec des tailles fixées. Soit \mathbf{p} une loi de reproduction. Pour tout entier n suffisamment grand tel que $\mathbb{P}(|\tau| = n) > 0$, on définit alors $\tilde{\tau}^{(n)}$ comme un arbre de Galton-Watson conditionné à avoir n nœuds, c'est à dire un arbre aléatoire de loi :

$$\mathbb{P}(\tilde{\tau}^{(n)} = \mathbf{t}) = \frac{\mathbb{P}(\tau = \mathbf{t})}{\mathbb{P}(|\tau| = n)}, \text{ pour tout } \mathbf{t} \in \mathbb{T} \text{ tel que } |\mathbf{t}| = n.$$

Par définition, la taille de $\tilde{\tau}^{(n)}$ est égale à n .

Nous allons maintenant définir les arbres simplement générés qui sont une généralisation des arbres de Galton-Watson conditionnés par leur taille totale. Ils ont été introduits par Meir et Moon [144] en 1978. On considère une suite de poids $\mathbf{q} = (\mathbf{q}(k), k \in \mathbb{N})$ de réels positifs telle que $\sum_{k \in \mathbb{N}} \mathbf{q}(k) > \mathbf{q}(1) + \mathbf{q}(0)$ et $\mathbf{q}(0) > 0$ (autrement dit, $\mathbf{q}(0) > 0$ et il existe $k \geq 2$ tel que $\mathbf{q}(k) > 0$). Ces hypothèses sont faites pour éviter les cas triviaux. Pour tout $\mathbf{t} \in \mathbb{T}_0$, on définit le poids $w(\mathbf{t})$ de \mathbf{t} par :

$$w(\mathbf{t}) = \prod_{v \in \mathbf{t}} \mathbf{q}(k_v(\mathbf{t})).$$

On pose pour $p \in \mathbb{N}^*$:

$$w(\mathbb{T}^{(p)}) = \sum_{\mathbf{t} \in \mathbb{T}^{(p)}} w(\mathbf{t}).$$

Définition. Pour $p \in \mathbb{N}^*$ tel que $w(\mathbb{T}^{(p)}) > 0$, un arbre simplement généré à valeurs dans $\mathbb{T}^{(p)}$ et associé à la suite de poids \mathbf{q} est une variable aléatoire $\tau^{(p)}$ à valeurs dans $\mathbb{T}^{(p)}$ dont la loi est caractérisée par :

$$\mathbb{P}(\tau^{(p)} = \mathbf{t}) = \frac{w(\mathbf{t})}{w(\mathbb{T}^{(p)})}, \quad \mathbf{t} \in \mathbb{T}^{(p)}.$$

En particulier, d'après (1.2), si la suite de poids \mathbf{q} est une loi de probabilité sur \mathbb{N} alors $\tau^{(p)}$ a la même loi qu'un arbre de Galton-Watson de loi de reproduction \mathbf{q} et conditionné à avoir p nœuds.

Même quand \mathbf{q} n'est pas une loi de probabilité, on peut sous certaines conditions sur la loi \mathbf{q} , voir les arbres simplement générés comme des arbres de Galton-Watson conditionnés, voir Janson [114], section 4. On dit que deux suites de poids $\mathbf{q} = (\mathbf{q}(k), k \in \mathbb{N})$ et $\tilde{\mathbf{q}} = (\tilde{\mathbf{q}}(k), k \in \mathbb{N})$ sont équivalentes s'ils existent $a, b > 0$ tels que :

$$\tilde{\mathbf{q}}(k) = ab^k \mathbf{q}(k), \quad \forall k \in \mathbb{N}.$$

Il est facile de voir que deux suites de poids équivalentes définissent les mêmes arbres simplement générés. On introduit la fonction génératrice $g_{\mathbf{q}}$ de \mathbf{q} : $g_{\mathbf{q}}(\theta) = \sum_{k \in \mathbb{N}} \theta^k \mathbf{q}(k)$ pour $\theta > 0$. Notons $\rho \in [0, +\infty]$ son rayon de convergence. Dans [114], section 4, Janson montre qu'il existe une loi de probabilité équivalente à la suite de poids \mathbf{q} si et seulement si $\rho > 0$. Dans ce cas, les lois de probabilité équivalentes à \mathbf{q} sont données par :

$$\mathbf{p}(k) = \frac{t^k \mathbf{q}(k)}{g_{\mathbf{q}}(t)}, \quad \text{pour tout } t > 0 \text{ tel que } g_{\mathbf{q}}(t) < \infty.$$

Ainsi, un arbre simplement généré dont la suite de poids \mathbf{q} vérifie la condition précédente peut-être défini par une loi de probabilité \mathbf{p} équivalente à \mathbf{q} et donc être vu comme un arbre de Galton-Watson conditionné. Ceci signifie que $\tilde{\tau}^{(\mathbf{p})}$ associée à \mathbf{p} et $\tau^{(\mathbf{p})}$ associé à \mathbf{q} ont même loi. Il est également possible dans certains cas de se ramener à une loi de probabilité critique. On dit que \mathbf{q} est générique si l'équation $sg'_{\mathbf{q}}(s) = g_{\mathbf{q}}(s)$ admet une unique solution $s_{\mathbf{q}}$. Dans ce cas, la loi de probabilité \mathbf{p} définie par

$$\mathbf{p}(k) = \frac{s_{\mathbf{q}}^k \mathbf{q}(k)}{g_{\mathbf{q}}(s_{\mathbf{q}})}, \quad k \geq 0,$$

est une suite de poids équivalente à \mathbf{q} . De plus, la probabilité \mathbf{p} est une loi de probabilité critique, voir Janson [114], section 4.

Plusieurs classes combinatoires d'arbres aléatoires à n nœuds peuvent être réalisées comme des arbres simplement générés à n nœuds et donc aussi comme des arbres de Galton-Watson conditionnés à avoir n nœuds, voir table 1.1. C'est le cas des arbres planaires à n nœuds, des arbres de Cayley (arbres enracinés, étiquetés et non ordonnés) à n nœuds et des arbres binaires complets à n nœuds à condition que n soit impair (et donc a $\frac{n-1}{2}$ nœuds internes), que l'on munit de la probabilité uniforme (i.e. dans chaque classe, chaque arbre est équiprobable), voir Aldous [9], Drmota [68], partie 1.2.7 ou Janson [114], chapitre 10.

Une des propriétés remarquables des arbres simplement générés est qu'ils ont tendance à avoir une structure déséquilibrée, i.e. ils contiennent quelques branches de grandes tailles. Ainsi, la hauteur moyenne d'un arbre simplement généré de taille n est d'ordre \sqrt{n} (par opposition à un arbre de taille n dit équilibré qui aura une hauteur d'ordre $\log(n)$) (voir [68], chapitres 3 et 4 ou Janson [114], section 21.3).

Remarquons ici que nous nous sommes intéressés aux arbres de Galton-Watson conditionnés par leur taille totale, mais d'autres conditionnements ont également été étudiés comme par exemple, avoir une grande hauteur ou avoir un grand nombre de feuilles. Par opposition aux limites d'échelle des arbres de Galton-Watson, qui est une convergence globale, on peut ainsi s'intéresser aux limites locales de ces arbres. Dans son papier fondateur [122] de 1986, Kesten a été le premier à s'intéresser aux limites locales d'un arbre de Galton-Watson critique ou sous-critique conditionné à avoir une grande hauteur. Il a montré que la limite est l'arbre biaisé par la taille avec une unique branche infinie. Pour d'autres références sur les limites locales d'arbres de Galton-Watson conditionnés avec des conditionnements plus généraux, voir Jonsson et Stefánsson [119], Janson [114], chapitre 7 ou encore Abraham et Delmas [4, 3].

1.2.3 Convergence du processus de contour des arbres de Galton-Watson conditionnés

On commence par définir l'excursion brownienne normalisée notée $B_{\text{ex}} = (B_{\text{ex}}(t), 0 \leq t \leq 1)$. De manière informelle, B_{ex} est un mouvement brownien standard partant de l'origine et conditionné à rester strictement positif sur $(0, 1)$ et à revenir en 0 au temps 1. Il existe plusieurs manières de définir et caractériser une excursion brownienne (voir Revuz et Yor [168], chapitre 12, partie 2 ou McKean [112], partie 2.9). Biane [23], Verwaat [189] ou encore

Classe d'arbres à n nœuds munie de la probabilité uniforme	Arbres planaires	Arbres de Cayley	Arbres binaires complets
Cardinal de l'ensemble	C_{n-1}	n^{n-1}	$C_{\frac{n-1}{2}}$
Suite de poids de l'arbre simplement généré	$\mathbf{p}(k) = 1, \forall k \geq 0$	$\mathbf{p}(k) = \frac{1}{k!}, \forall k \geq 0$	$\mathbf{p}(0) = \mathbf{p}(2) = 1$ et $\mathbf{p}(k) = 0, \forall k \neq 0, 2$
Fonction génératrice associée à la suite de poids	$g_{\mathbf{p}}(t) = \frac{1}{1-t}$	$g_{\mathbf{p}}(t) = e^t$	$g_{\mathbf{p}}(t) = 1 + t^2$
Loi critique équivalente π	Loi géométrique de paramètre $\frac{1}{2}$	Loi de Poisson de paramètre 1	$\pi(0) = \frac{1}{2}$ et $\pi(2) = \frac{1}{2}$
Variance σ^2 de π	$\sigma^2 = 2$	$\sigma^2 = 1$	$\sigma^2 = 1$

TABLE 1.1 – Exemples d'arbres aléatoires vus comme des arbres de Galton-Watson conditionnés.

Rogers et Williams [171], partie 40 ont montré que l'excursion brownienne normalisée peut être construite à partir d'un pont brownien (processus stochastique dont la loi est celle d'un mouvement brownien standard (i.e. issu de 0) et conditionné à valoir 0 au temps 1), lui-même pouvant être décrit en fonction d'un mouvement brownien standard. En fait, l'excursion brownienne normalisée a même loi qu'un pont de Bessel de dimension 3 i.e. que la norme d'un pont brownien de dimension 3. Ainsi, pour représenter une trajectoire d'une excursion brownienne normalisée, on construit un pont brownien de dimension 3 noté $X(t) = (X_1(t), X_2(t), X_3(t))$ à partir d'un mouvement brownien de dimension 3, $B(t) = (B_1(t), B_2(t), B_3(t))$ de la manière suivante :

$$X_i(t) = B_i(t) - tB_i(1), \text{ pour tout } t \in [0, 1], i \in \{1, 2, 3\}.$$

Enfin, l'excursion brownienne normalisée B_{ex} est obtenue de la manière suivante :

$$B_{\text{ex}}(t) = \|X(t)\| = \sqrt{X_1(t)^2 + X_2(t)^2 + X_3(t)^2}, \text{ pour tout } t \in [0, 1].$$

On note $\mathbb{N}^{(1)}$ la loi de l'excursion brownienne normalisée.

Soit \mathbf{p} une loi de reproduction sur \mathbb{N} critique (i.e. $\sum_{k \in \mathbb{N}} k\mathbf{p}(k) = 1$). On note $\sigma^2 > 0$ sa variance finie ou infinie. Soit τ un arbre de Galton-Watson de loi de reproduction \mathbf{p} et pour chaque p tel que $\mathbb{P}(|\tau| = p) > 0$, soit $\tau^{(p)}$ un arbre de même loi que τ conditionnellement à $\{|\tau| = p\}$. Nous allons donner les résultats de convergence du processus de contour de $\tau^{(p)}$.

La variance de \mathbf{p} est finie : Dans le cas où la variance σ^2 est finie, Aldous [10] a montré que le processus de contour $C^{\tau^{(p)}}$ de $\tau^{(p)}$ correctement normalisé converge en loi vers l'excursion brownienne normalisée.

Theorem (Aldous). *On suppose que la loi de reproduction \mathbf{p} est critique et que $0 < \sigma^2 < +\infty$. Alors on a la convergence en loi dans l'espace $\mathcal{C}([0, 1])$ muni de la convergence uniforme :*

$$\frac{\sigma}{2\sqrt{p}} \left(C^{\tau^{(p)}}(2ps), s \in [0, 1] \right) \xrightarrow[p \rightarrow +\infty]{(d)} (B_{ex}(s), s \in [0, 1]),$$

où la convergence a lieu pour la sous-suite infinie de p tels que $\mathbb{P}(|\tau| = p) > 0$.

La loi \mathbf{p} est dans le domaine d'attraction d'une loi stable : Une généralisation du théorème d'Aldous dans la cas où \mathbf{p} appartient au domaine d'attraction d'une loi stable a été obtenue par Duquesne [69]. On suppose donc que \mathbf{p} est dans le domaine d'attraction d'une loi stable. Rappelons la définition. Soit $(U_k, k \in \mathbb{N}^*)$ une suite de variables indépendantes de même loi de distribution \mathbf{p} . On pose $W_p = \sum_{k=1}^p U_k - p$. On dit que \mathbf{p} est dans le domaine d'attraction d'une loi stable d'exposant de Laplace $\psi(\lambda) = \kappa\lambda^\gamma$ avec $\gamma \in (1, 2]$ et $\kappa > 0$, et de suite de normalisation $(a_p, p \in \mathbb{N}^*)$ de réels positifs ou nuls, si W_p/a_p converge en loi, quand p tend vers l'infini, vers une variable aléatoire X d'exposant de Laplace ψ (i.e. on a $\mathbb{E}[e^{-\lambda X}] = e^{-\psi(\lambda)}$ pour $\lambda \geq 0$).

Remarque. Si \mathbf{p} a une variance σ^2 finie alors on peut choisir $a_p = \sqrt{p}$ et X est alors de loi normale centrée de variance σ^2 , de telle sorte que $\psi(\lambda) = \sigma^2\lambda^2/2$.

Theorem (Duquesne). *On suppose que la loi de reproduction \mathbf{p} est critique et qu'elle est dans le domaine d'attraction d'une loi stable d'exposant $\psi(\lambda) = \kappa\lambda^\gamma$ avec $\gamma \in (1, 2]$ et $\kappa > 0$, et de suite de normalisation $(a_p, p \in \mathbb{N}^*)$. Alors il existe un processus aléatoire non trivial, continu, positif ou nul $H = (H_s, s \in [0, 1])$, tel que on ait la convergence en loi dans l'espace $\mathcal{C}([0, 1])$ muni de la convergence uniforme :*

$$\frac{a_p}{p} \left(C^{\tau^{(p)}}(2ps), s \in [0, 1] \right) \xrightarrow[p \rightarrow +\infty]{(d)} H,$$

où la convergence a lieu pour la sous-suite infinie de p tels que $\mathbb{P}(|\tau| = p) > 0$.

Le processus H , voir Duquesne [69] pour une construction de H , désigne une excursion normalisée du processus hauteur, introduit dans Duquesne et Le Gall [70], d'un arbre de Lévy de mécanisme de branchement ψ .

1.3 Arbres réels

Dans cette partie, on introduit la notion d'arbres réels, définis en tant qu'espaces métriques. Nous verrons que ces arbres peuvent être codés par une fonction continue de la même manière que les arbres discrets peuvent l'être avec leur fonction de contour.

1.3.1 Définitions

Nous allons définir les arbres réels ou \mathbb{R} -arbres en terme d'espaces métriques (voir Evans [78]) :

Définition. *Un arbre réel (resp. arbre réel compact) est un espace métrique (resp. espace métrique compact) (\mathcal{T}, d) qui satisfait les deux propriétés suivantes pour tout $x, y \in \mathcal{T}$:*

- *Il existe une unique isométrie $f_{x,y}$ de $[0, d(x, y)]$ dans \mathcal{T} telle que l'on ait $f_{x,y}(0) = x$ et $f_{x,y}(d(x, y)) = y$.*
- *Si ϕ est une application continue injective de $[0, 1]$ dans \mathcal{T} telle que $\phi(0) = x$ et $\phi(1) = y$, alors on a $\phi([0, 1]) = f_{x,y}([0, d(x, y)])$.*

Un arbre réel enraciné est un arbre réel (\mathcal{T}, d) ayant un point distingué \emptyset appelé la racine.

De manière équivalente, un espace métrique (\mathcal{T}, d) est un arbre réel si et seulement si :

- \mathcal{T} est connecté
- d satisfait la condition suivante appelée « condition des 4 points » :

$$d(s, t) + d(x, y) \leq \max(d(s, x) + d(t, y), d(s, y) + d(t, x)) \quad \text{pour tout } s, t, x, y \in \mathcal{T}.$$

On peut munir l'espace des arbres réels enracinés et compacts d'une distance appelée distance de Gromov-Hausdorff qui rend cet espace polonais, voir par exemple Le Gall [135], théorème 2.1.

De la même manière que la fonction de contour code les planaires, on peut coder certains arbres réels par une fonction continue. Soit h une fonction continue de $[0, 1]$ dans \mathbb{R}^+ telle que $h(0) = h(1) = 0$. Pour $x, y \geq 0$. On définit

$$d_h(x, y) = h(x) + h(y) - 2 \min_{z \in [x \wedge y, x \vee y]} h(z),$$

où $x \wedge y = \min(x, y)$ et $x \vee y = \max(x, y)$. Il est facile de vérifier que d_h est symétrique et satisfait l'inégalité triangulaire et donc que c'est une pseudo-distance sur $[0, 1]$. On introduit une relation d'équivalence \sim_h définie sur $[0, 1]^2$ par $x \sim_h y \Leftrightarrow d_h(x, y) = 0$. L'arbre continu codé par h est l'espace métrique quotient $\mathcal{T}_h = [0, 1] / \sim_h$ muni de la distance d_h et enraciné en la classe d'équivalence de 0. Il est facile de voir que (\mathcal{T}_h, d_h) est un arbre réel compact (voir Le Gall et Miermont [137], Théorème 3.1). On note \mathbf{p}_h la projection canonique de $[0, 1]$ dans \mathcal{T}_h . Pour tout $t \in [0, 1]$, $\mathbf{p}_h(t)$ est donc un nœud de \mathcal{T}_h à distance $h(t)$ de la racine. Pour $0 \leq s \leq t \leq 1$, l'ancêtre commun aux nœuds $\mathbf{p}_h(s)$ et $\mathbf{p}_h(t)$ est à distance $\min_{s \leq r \leq t} h(r)$ de la racine. La figure 1.4 représente le sous-arbre de l'arbre \mathcal{T}_h engendré par les feuilles $\mathbf{p}_h(s), \mathbf{p}_h(t), \mathbf{p}_h(u)$ et la racine.

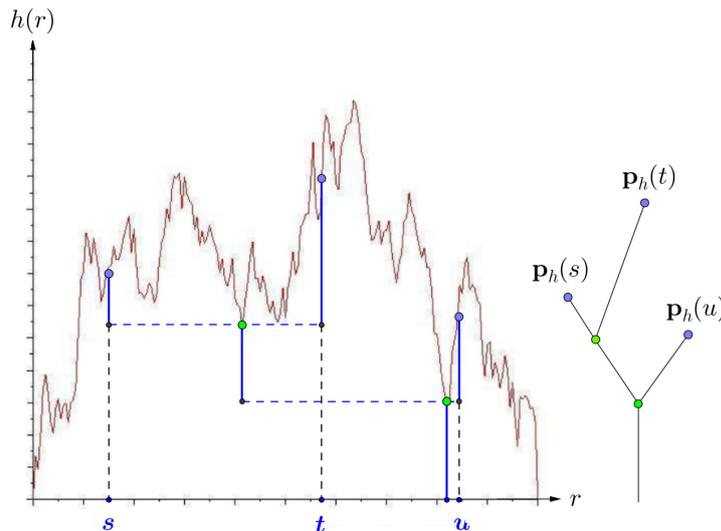


FIGURE 1.4 – Un sous-arbre de l'arbre \mathcal{T}_h

En fait, il est possible de montrer que tout arbre réel compact peut être codé par une fonction continue (voir Le Gall [135], partie 2.2).

1.3.2 Arbre brownien

Dans une série de papiers [8, 9, 10] datant du début des années 1990, Aldous introduit et développe la notion d'arbre continu brownien (d'abord appelé arbre continu compact) ainsi que plusieurs résultats de convergence. On rappelle que $B_{\text{ex}} = (B_{\text{ex}}(t) : 0 \leq t \leq 1)$ désigne l'excursion brownienne normalisée.

Définition. *L'arbre continu brownien (abrégé en CRT pour « continuum random tree » en anglais) est l'arbre aléatoire réel $\mathcal{T}_{B_{\text{ex}}}$ codé par l'excursion brownienne normalisée B_{ex} .*

Cette définition a été donnée par Aldous dans [10], corollaire 22 avec une différence d'un facteur 2 i.e. le CRT est codé par $2B_{\text{ex}}$ et non B_{ex} . Ainsi, afin d'énoncer nos résultats dans les deux cas précédents, il sera utile d'introduire le processus $e = \sqrt{2/\alpha}B_{\text{ex}}$ avec $\alpha > 0$. En effet, pour $\alpha = 1/2$, on est dans le cadre de l'arbre brownien défini par Aldous et pour $\alpha = 2$, on est dans le cadre de l'arbre brownien associé à l'excursion brownienne normalisée B_{ex} .

Comme e est un processus stochastique, \mathcal{T}_e est une variable aléatoire dans l'espace polonais des arbres aléatoires réels enracinés et compacts. De plus, puisque les minima locaux du mouvement brownien (et donc de l'excursion brownienne e) sont toujours distincts presque sûrement, \mathcal{T}_e est un arbre binaire presque sûrement (i.e. les nœuds ont 0 ou 2 enfants).

Dans [8, 9, 10], Aldous énonce quatre autres définitions équivalentes pour construire l'arbre brownien réel :

- par séparation poissonnienne d'une droite (« Poisson line-breaking » en anglais), voir Aldous [8] ou Pitman [163], partie 7.4.
- comme limite d'échelle d'arbres de Galton-Watson ayant une loi limite critique et de variance finie, conditionnés à avoir un grand nombre de nœuds, voir Aldous [9] et Le Gall [135].
- comme limite d'échelle d'arbres non étiquetés et non enracinés au sens de Gromov-Hausdorff (conjecture d'Aldous [9] p55, et prouvé en 2014 par Stufler [181]).
- par ses marginales finies dimensionnelles i.e. la loi de ses sous-arbres engendrés/couverts (« spanning trees » en anglais) par des feuilles choisies uniformément au hasard (voir Aldous [10], partie 4.3). Cette dernière construction sera développée dans la section suivante.

1.3.3 Construction d'arbres binaires à partir de l'arbre continu brownien

Dans cette partie, on considère des arbres marqués $\tilde{\mathbf{t}} = (\mathbf{t}, (h_v, v \in \mathbf{t}))$ où $h_v \geq 0$, pour tout $v \in \mathbf{t}$. h_v est la longueur de la branche en dessous de v et \mathbf{t} est le squelette de l'arbre marqué $\tilde{\mathbf{t}}$, voir partie 1.1.2. On note $\tilde{\mathbb{T}}^{(p)}$ désigne l'ensemble des arbres marqués à p feuilles. Si $\tilde{\mathbf{t}} = (\mathbf{t}, (h_v, v \in \mathbf{t}))$ est un arbre marqué, la longueur de $\tilde{\mathbf{t}}$ est définie par :

$$L(\tilde{\mathbf{t}}) = \sum_{v \in \mathbf{t}} h_v.$$

Soient $n \in \mathbb{N}$ et $0 < t_1 < \dots < t_{n+1} < 1$. On considère le sous-arbre de l'arbre continu brownien τ_e engendré par les $n + 1$ feuilles $\mathbf{p}_e(t_1), \dots, \mathbf{p}_e(t_{n+1})$ et la racine \emptyset . On le note $\tau_e(t_1, \dots, t_{n+1})$. C'est un arbre binaire, enraciné, complet et ordonné à $n + 1$ feuilles et $2n + 1$ arêtes (n'oublions pas que la racine possède également une arête, voir figure 1.4). On peut lui associer un arbre marqué

$$\tilde{\mathbf{t}}(e; t_1, \dots, t_{n+1}) = (\mathbf{t}, \{h_v(e; t_1, \dots, t_{n+1}), v \in \mathbf{t}\}),$$

où intuitivement, $\mathbf{t} = \mathbf{t}(e; t_1, \dots, t_{n+1})$ est identique à $\tau_e(t_1, \dots, t_{n+1})$ mais avec des longueurs de branches égales à 1 et n'ayant pas de branche sous la racine. Pour tout $v \in \mathbf{t}$,

$h_v(e; t_1, \dots, t_{n+1})$ est la longueur de la branche dans $\tau_e(t_1, \dots, t_{n+1})$ sous le nœud correspondant à $v \in \mathbf{t}$. Ainsi, l'arbre squelette \mathbf{t} est un arbre discret complet, ordonné et enraciné à $n + 1$ feuilles et $2n$ arêtes. Pour une construction plus rigoureuse, voir Aldous [10], Le Gall [134, 135] ou encore Duquesne et Le Gall [70].

Un processus stochastique peut être décrit par ses lois finies dimensionnelles. Inspirés par cette idée, nous allons caractériser l'arbre continu brownien en terme de ses marginales finies dimensionnelles qui sont ses sous-arbres couverts par des feuilles choisies uniformément au hasard.

On définit la mesure uniforme Λ_n sur l'ensemble des arbres marqués dont le squelette appartient à l'ensemble $\mathbb{T}_{\text{bin}}^{(n)}$ des arbres binaires complets et ordonnés à n nœuds internes (et donc $n + 1$ feuilles) par :

$$\int \Lambda_n(d\tilde{\mathbf{t}})F(\tilde{\mathbf{t}}) = \sum_{\mathbf{t} \in \mathbb{T}_{\text{bin}}^{(n)}} \int \prod_{v \in \mathbf{t}} dh_v F(\mathbf{t}, \{h_v, v \in \mathbf{t}\}),$$

pour toute fonction F positive, bornée et continue sur l'ensemble des arbres marqués à n feuilles.

Le théorème suivant a d'abord été énoncé par Aldous [9], partie 2.4. On peut également trouver l'énoncé de ce résultat dans Le Gall [134] et [135] partie 2.6 et Pitman [163], chapitre 7. On rappelle que $\mathbb{N}^{(1)}$ est la loi de l'excursion brownienne normalisée.

Theorem (Aldous). *La loi de l'arbre $\tilde{\mathbf{t}} = \tilde{\mathbf{t}}(e; t_1, \dots, t_{n+1})$ sous la mesure de probabilité*

$$(n + 1)! \mathbf{1}_{\{0 \leq t_1 \leq \dots \leq t_{n+1} \leq 1\}} dt_1 \dots dt_n \mathbb{N}^{(1)}(de)$$

est

$$(n + 1)! 2 \alpha^{n+1} L(\tilde{\mathbf{t}}) \exp(-2L(\tilde{\mathbf{t}})^2) \Lambda_n(d\tilde{\mathbf{t}}).$$

Autrement dit, pour toute fonction F positive bornée et continue sur l'ensemble des arbres marqués à $n + 1$ feuilles, on a :

$$\begin{aligned} \int \mathbb{N}^{(1)}(de) \int_{0 \leq t_1 \leq \dots \leq t_{n+1} \leq 1} F(\tilde{\mathbf{t}}(e; t_1, \dots, t_{n+1})) \\ = 2 \alpha^{n+1} \int \Lambda_n(d\tilde{\mathbf{t}}) L(\tilde{\mathbf{t}}) \exp(-2L(\tilde{\mathbf{t}})^2) F(\tilde{\mathbf{t}}). \end{aligned}$$

De ce théorème, il est facile de voir que le squelette $\mathbf{t}(e; t_1, \dots, t_{n+1})$ de $\tilde{\mathbf{t}}(e; t_1, \dots, t_{n+1})$ est uniformément distribué sur l'ensemble des arbres binaires complets ordonnés à $n + 1$ feuilles dont le cardinal est donné par le n -ième nombre de Catalan : $C_n = (2n)! / (n!(n + 1)!)$. Ainsi, conditionnellement au squelette, sous la mesure de probabilité

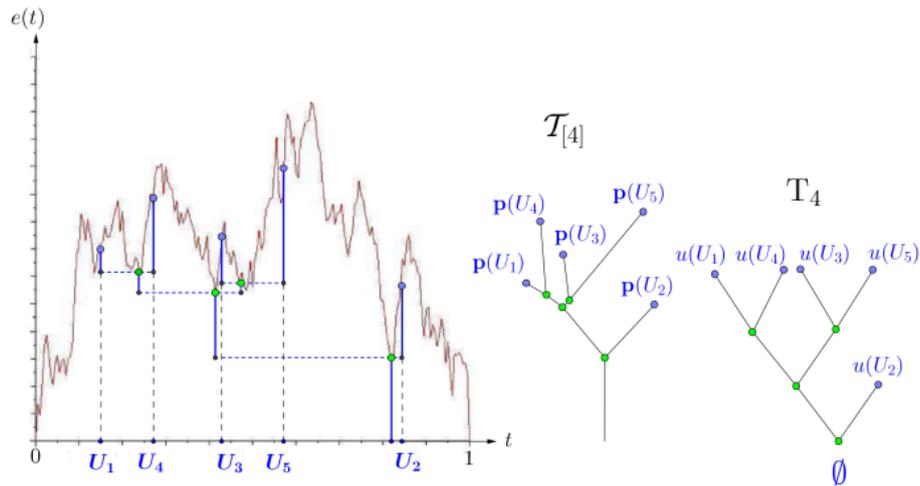
$$(n + 1)! \mathbf{1}_{\{0 \leq t_1 \leq \dots \leq t_{n+1} \leq 1\}} dt_1 \dots dt_n \mathbb{N}^{(1)}(de),$$

la densité du vecteur $(h_v(e; t_1, \dots, t_{n+1}), v \in \mathbf{t})$ est donnée par :

$$C_n 2 \alpha^{n+1} (n + 1)! L(\tilde{\mathbf{t}}) \exp(-2L(\tilde{\mathbf{t}})^2) = 2 \frac{(2n)!}{n!} \alpha^{n+1} L(\tilde{\mathbf{t}}) \exp(-2L(\tilde{\mathbf{t}})^2).$$

Les longueurs d'arêtes sont des variables échangeables et sont indépendantes du squelette.

On se donne maintenant $(U_n, n \in \mathbb{N}^*)$ une suite de variable aléatoires indépendantes de loi uniforme sur $[0, 1]$ et indépendante de e . On note $\mathcal{T}_{[n]}$ l'arbre aléatoire réel engendré par les $n + 1$ feuilles $\mathbf{p}(U_1), \dots, \mathbf{p}(U_{n+1})$ et la racine et $\tilde{\mathbf{T}}_n = (\mathbf{T}_n; (h_{n,v}, v \in \mathbf{T}_n))$ l'arbre marqué

FIGURE 1.5 – L'excursion brownienne, \tilde{T}_n et T_n (pour $n = 4$).

associé (voir figure 1.5 pour une simulation avec $n = 4$). On note L_n la longueur de l'arbre marqué \tilde{T}_n .

Le théorème précédent se réécrit ainsi de la manière suivante : la loi de probabilité de \tilde{T}_n est

$$(n+1)! 2 \alpha^{n+1} L_n \exp(-2L_n^2) \Lambda_{n+1}(d\tilde{T}_n),$$

et la densité de $(h_{n,v}, v \in T_n)$ est, conditionnellement à T_n , donnée par :

$$f_n((h_{n,v}, v \in T_n)) = 2 \frac{(2n)!}{n!} \alpha^{n+1} L_n e^{-\alpha L_n^2} \prod_{v \in T_n} \mathbf{1}_{\{h_{n,v} > 0\}}. \quad (1.3)$$

En particulier, L_n est indépendante de T_n et a même loi que $\sqrt{\Delta_n/\alpha}$ où Δ_n est une loi Gamma de paramètre $n+1$ et 1 (voir Aldous [10] ou Pitman [163], Théorème 7.9). Ainsi, conditionnellement à T_n , la densité de L_n est donnée par :

$$f_{L_n}(x) = 2 \frac{\alpha^{n+1}}{n!} x^{2n+1} e^{-\alpha x^2} \mathbf{1}_{\{x > 0\}}. \quad (1.4)$$

1.4 Fonctionnelles de coût et fonctions péage

Les fonctionnelles de coût sur les arbres planaires ont été largement étudiées dans des domaines très variés (informatique, physique, chimie, biologie ...), voir les références dans les parties 1.4.2 et 1.4.3. Nous allons étudier l'asymptotique de ces fonctionnelles définies sur les arbres planaires (arbres enracinés et ordonnés) quand le nombre de nœuds tend vers l'infini. Les fonctionnelles que nous allons considérer sont additives (définies par une relation de récurrence) et induites par une fonction dite péage. La valeur d'une fonctionnelle additive sur un arbre est définie de manière récursive comme la somme de la valeur de la fonctionnelle sur les sous-arbres enracinés sur les enfants de la racine de l'arbre à laquelle on ajoute un terme (le péage) qui est une fonction qui dépend de la taille de l'arbre. Les fonctionnelles de coût définies de la sorte représentent le coût des algorithmes du type diviser et conquérir, où la nature récursive intrinsèque de ces algorithmes amène naturellement à une telle formulation.

1.4.1 Définition

Définition. Une fonctionnelle de coût F sur les arbres planaires est une fonctionnelle additive sur les arbres si elle satisfait la relation de récurrence suivante :

$$F(\mathbf{t}) = \sum_{i=1}^{k_\emptyset(\mathbf{t})} F(\mathbf{t}_i) + b_{|\mathbf{t}|}$$

pour tout arbre $\mathbf{t} \in \mathbb{T}_0$ tel que $|\mathbf{t}| \geq 1$ et avec $F(\emptyset) = 0$ et où $\mathbf{t}_1, \dots, \mathbf{t}_{k_\emptyset(\mathbf{t})}$ sont les sous-arbres enracinés aux enfants de la racine de \mathbf{t} . La suite $(b_k, k \geq 1)$ est appelée la fonction péage.

Par définition de F , il est facile de voir que l'on a pour tout $\mathbf{t} \in \mathbb{T}_0$:

$$F(\mathbf{t}) = \sum_{v \in \mathbf{t}} b_{|\mathbf{t}_v|}. \quad (1.5)$$

Remarque. 1. En particulier, une fonctionnelle de coût F sur les arbres binaires est une fonctionnelle additive sur les arbres si elle satisfait la relation de récurrence suivante :

$$F(\mathbf{t}) = F(L(\mathbf{t})) + F(R(\mathbf{t})) + b_{|\mathbf{t}|}$$

pour tout arbre \mathbf{t} tel que $|\mathbf{t}| \geq 1$ et avec $F(\emptyset) = 0$ et où $L(\mathbf{t}) = \mathbf{t}_1$ (resp. $R(\mathbf{t}) = \mathbf{t}_2$) désigne le sous-arbre de gauche (resp. de droite) de la racine de \mathbf{t} .

2. La fonction péage peut également dépendre de manière plus générale de l'arbre lui-même. Nos résultats concerneront le cas où elle ne dépend que du cardinal de l'arbre, voir Flajolet, Gourdon et Martínez [92], Devroye [57] et Wagner [190].

1.4.2 Transition de phase

Par la suite, on distinguera les fonctions de coût dites locales et globales. Dans le chapitre 3, nous nous restreindrons à l'étude de fonctionnelles globales. Ces deux termes sont utilisés pour mettre en évidence la transition de phase entre le régime normal et non-normal des lois limites. Intuitivement, lorsque l'on considère des fonctions de coût dites locales, i.e. des fonctions de coût petites, la somme est dominée par la contribution des très nombreux petits sous-arbres. Comme les différentes parties des arbres ont une très faible dépendance les unes entre les autres, cela rend possible la normalité asymptotique. Pour une fonction péage dite globale, i.e. une fonction qui croît très rapidement avec la taille de l'arbre, la somme sera au contraire dominée par la contribution des très grands arbres, qui sont plus fortement dépendants les uns des autres, faisant donc apparaître d'autres distributions limites.

Quand la fonction de coût dépend seulement de la taille de l'arbre, la transition de phase entre les différents régimes pour les arbres binaires a été étudiée par Hwang et Neininger [111] pour le modèle de permutations uniformes et par Fill et Kapur [89] pour le modèle de Catalan.

Dans le cas où les fonctions de coût dépendent de l'arbre lui-même Wagner [191] étudie la transition de phase pour des arbres simplement générés, des arbres récursifs et des arbres binaires de recherche. Le fait que les fonctions de coût peuvent dépendre de l'arbre lui-même amplifie les relations de dépendance rendant difficile la détermination de la zone de transition de phase.

Nous allons consacrer la fin de ce paragraphe aux fonctionnelles locales, qui bien que n'entrant pas dans notre cadre d'étude, ont été très étudiées. Nous en donnons quelques exemples : taille totale ($b_k = 1$), nombre de feuilles ($b_k = \mathbf{1}_{\{k=1\}}$), nombre de nœuds protégés (un nœud est dit protégé si ce n'est pas une feuille ni le parent d'une feuille), nombre de sous-arbres de taille donnée p ($b_k = \mathbf{1}_{\{k=p\}}$) et nombre total de sous-arbres, voir Janson [115], partie 2.

On peut également considérer des fonctions péage qui dépendent plus généralement de l'arbre $\mathbf{t} \in \mathbb{T}$ et non plus simplement de sa taille $|\mathbf{t}|$: par exemple, on peut compter le nombre de sous-arbres précisément égaux à un sous-arbre donné \mathbf{t}_0 (« tree patterns » en anglais) dont la fonction péage est donnée par $\mathbf{1}_{\{\mathbf{t}=\mathbf{t}_0\}}$, voir Flajolet, Gourdon et Martínez [92], Devroye [57] et Wagner [190].

La fonctionnelle de forme (« shape functional » en anglais) $Q(\mathbf{t})$, définie dans (1.1), est également une fonctionnelle qui a suscité un grand intérêt. Elle a d'abord été étudiée dans le cadre des arbres binaires de recherche sous le modèle de permutations uniformes, voir Dobrow et Fill [64] ou Fill [85]. Elle est appelée fonctionnelle de forme car elle permet d'évaluer la forme de l'arbre \mathbf{t} puisque les valeurs maximales de $Q(\mathbf{t})$ sont atteintes pour les arbres qui sont les plus équilibrés. La fonctionnelle additive $-\log Q$ est associée à la fonction péage $b_k = \log(k)$. Sa convergence a été étudiée par Fill [85, 89] pour les arbres binaires sous le modèle de Catalan et de permutations uniformes et Meir et Moon [145] pour les arbres simplement générés. Cette fonctionnelle intervient aussi dans une méthode d'auto-organisation des arbres binaires de recherche appelée « move-to-root », voir Dobrow et Fill [62, 63].

De nombreux articles fournissent des résultats asymptotiques pour des fonctionnelles additives locales assez générales. Ces études permettent notamment de retrouver le comportement limite des fonctionnelles citées précédemment. Le lecteur pourra se référer aux articles suivants : pour les arbres binaires de recherche sous le modèle de permutations uniformes, voir Devroye [60], Holmgren et Janson [108] et Wagner [191], pour les arbres de Galton-Watson conditionnés (i.e. les arbres simplement générés), voir Janson [115] et Wagner [191], pour les arbres récursifs, voir Holmgren et Janson [108] et Wagner [191], pour les arbres étiquetés, voir Wagner [190], pour les arbres m -aire de recherche, voir Fill et Kapur [90], et enfin pour les arbres m -aire croissants de recherche, voir Ralaivaosaona et Wagner [165].

1.4.3 Fonctionnelles de coût globales

Nous donnons maintenant différents exemples de fonctionnelles additives globales que l'on peut trouver dans la littérature et qui interviennent naturellement en informatique, physique ou biologie. Nos résultats plus généraux permettent d'obtenir des asymptotiques sur des fonctionnelles additives globales plus générales.

On note d la distance usuelle de graphe sur $\mathbf{t} \in \mathbb{T}$. Ainsi, pour $v, w \in \mathbf{t}$, w est un ancêtre de v (que l'on écrit $w \preceq v$, voir partie 1.1.1) si $d(\emptyset, v) = d(\emptyset, w) + d(w, v)$. Pour $u, v \in \mathbf{t}$, on note $u \wedge v$, l'ancêtre le plus récent de u et v : $u \wedge v$ est l'unique élément de \mathbf{t} tel que : $w \preceq u$ et $w \preceq v$ implique que $w \preceq u \wedge v$.

Pour $\mathbf{t} \in \mathbb{T}$, on a par exemple les fonctionnelles suivantes :

- **Longueur de cheminement total :**

$$P(\mathbf{t}) = \sum_{v \in \mathbf{t}} d(\emptyset, v) = \sum_{v \in \mathbf{t}} |\mathbf{t}_v| - |\mathbf{t}|.$$

La fonction péage est donnée par : $b_k = k - 1$. Dans les cas des arbres binaires de recherche, la fonctionnelle compte le nombre de comparaisons de l'algorithme de tri rapide (« quicksort » en anglais) pour trier une liste de nombre distincts, voir Rösler [172].

Pour les arbres binaires sous le modèle de Catalan et les arbres de Galton-Watson conditionnés à avoir un certain nombre de nœuds, il est connu que cette fonctionnelle converge vers la loi de Airy, $2 \int_0^1 B_s ds$ où B est un mouvement brownien standard sur

$[0, 1]$, voir Takács [183], Aldous [9, 10] et Janson [113]. L'asymptotique de cette fonctionnelle a également été étudiée pour les arbres binaires sous le modèle de permutations uniformes, voir Régnier [166], Rösler [172] et Fill et Janson [87] et pour les arbres m -aire de recherche, voir Fill et Kapur [90, 91].

- **Indice de Wiener :**

$$W(\mathbf{t}) = \sum_{u,v \in \mathbf{t}} d(u,v) = 2|T| \sum_{w \in T} |T_w| - 2 \sum_{w \in T} |T_w|^2.$$

La fonctionnelle fait intervenir les fonctions de péage $b_k = k$ et $b_k = k^2$. L'indice de Wiener a été introduit en 1947 par Harold Wiener [193] comme le nombre de chemins (« path number » en anglais). Il était initialement défini comme le nombre de liens entre chaque paire d'atomes dans une molécule acyclique. L'indice de Wiener joue un rôle important dans les propriétés physiques-chimiques des structures chimiques (point d'ébullition, défauts de cristaux ...), voir [101, 66, 185].

La convergence a été étudiée par Janson [113] et Chassaing et Janson [116] pour les arbres binaires sous le modèle de Catalan, Neininger [152] pour les arbres binaires de recherche et les arbres récursifs et Janson [113] pour les arbres de Galton-Watson conditionnés à avoir un certain nombre de nœuds.

- **Fonctionnelles de coût associées aux fonctions péages de type puissance :**
 $b_k = k^\beta$ avec $\beta > 0$.

Fill et Kapur [89] ont montré la convergence en loi de la fonctionnelle pour les arbres binaires sous le modèle de Catalan. La limite est caractérisée en termes de ses moments. Leurs preuves reposent sur des calculs combinatoires. Plus tard, Fill et Janson [88] conjecturent une expression de la limite qui s'exprime en fonction de l'excursion normalisée. Le cas des arbres binaires de recherche a été traité par Neininger [151] (avec $\beta \in \mathbb{R}$, $\beta > 1$) et les arbres m -aire par Fill et Kapur [90, 91].

De nombreuses fonctionnelles globales étudiées dans la littérature se réécrivent en termes de fonctionnelles associées aux fonctions péages $b_k = k^\beta$ avec $\beta > 0$.

Fonctionnelles de coût sur les arbres phylogénétiques

De nombreux exemples de fonctionnelles proviennent également de la biologie au travers des arbres phylogénétiques qui sont utiles pour classer des populations et représenter l'évolution des relations entre les espèces (voir Aldous [13], Ford [93]).

Un arbre phylogénétique de taille n (aussi appelé « cladogram » en anglais) est un arbre binaire complet enraciné à n feuilles étiquetées de 1 à n qui correspondent aux espèces et à $n - 1$ nœuds internes qui correspondent à leurs ancêtres, voir Ford [93]. Les longueurs de branches ne seront pas prises en compte, si bien que l'on s'intéressera seulement à la topologie des arbres i.e. à leur forme.

On peut munir les arbres phylogénétiques de deux modèles de probabilité : le modèle de Yule-Harding ou modèle d'évolution neutre, défini par Yule [195] en 1924 (qui correspond exactement au modèle de permutations uniformes sur les ABR puisque un arbre phylogénétique à n feuilles peut être vu comme un ABR à $n - 1$ nœuds, voir Aldous [11]) et le modèle uniforme (i.e. le modèle de Catalan en informatique).

Quand le nombre d'espèces augmente, il est intéressant de regarder la forme des arbres afin de mieux comprendre leur structure. Par exemple, une asymétrie dans un arbre peut traduire une adaptation meilleure d'une ou plusieurs espèces. Nous donnons trois exemples d'indices sur les arbres phylogénétiques qui rentrent dans le cadre des fonctionnelles globales (voir Shao et Sokal [179], Heard [104], Kirkpatrick et Slatkin [123], Mooers et Heard [149], Felsenstein

[82], chapitre 33 pour d'autres exemples d'indice). L'intérêt de ces indices est qu'ils dépendent seulement de la forme des arbres et en particulier, ils sont invariants par isomorphisme et par renumérotation des feuilles. Les indices de Colless et cophénétique ne concernent que les arbres binaires.

- **Indice de Sackin (= longueur du chemin extérieur) :**

$$S(\mathbf{t}) = \sum_{v \in \mathcal{L}(\mathbf{t})} d(\emptyset, v),$$

où $\mathcal{L}(\mathbf{t})$ désigne l'ensemble des feuilles de l'arbre \mathbf{t} . L'indice de Sackin a été introduit en 1972 par Sackin [175].

Sa convergence a été étudiée par Blum, François et Janson [27] pour les arbres binaires pour le modèle de Catalan et le modèle de permutations uniformes. L'indice de Sackin a également été étudié en informatique dans le cadre des arbres binaires de recherche sous le modèle de permutations uniformes (donc sous le modèle de Yule), voir Régnier [166], Rösler [172]. Dans ce cas, cet indice est plus connu sous le nom de longueur de chemin extérieur (« external path length » en anglais) qui correspond au nombre de comparaisons dans l'algorithme quicksort.

- **Indice de Colless :**

$$C(\mathbf{t}) = \sum_{v \in T} \left| |\mathcal{L}(L_v)| - |\mathcal{L}(R_v)| \right|,$$

où $|\mathcal{L}(L_v)|$ (resp. $|\mathcal{L}(R_v)|$) est le nombre de feuilles du sous-arbre de gauche (resp. de droite) au dessus du nœud v . L'indice de Colless a quant à lui été introduit un peu plus tard, en 1982 par Colless [50], paragraphe 2.2, pour mesurer l'asymétrie d'un arbre binaire.

Sa convergence a été étudiée par Blum, François et Janson [27] pour le modèle de Catalan et le modèle de permutations uniformes.

- **Indice cophénétique :**

$$Co(\mathbf{t}) = \sum_{u, v \in \mathcal{L}(\mathbf{t}), u \neq v} d(\emptyset, u \wedge v).$$

Cet indice a été considéré récemment, en 2013, par Mir, Rosseló et Rotger [146] et Cardona, Mir et Rosseló [42] qui ont étudié certaines de ses propriétés asymptotiques sous les deux modèles, telles que son espérance et sa variance asymptotiques.

1.5 Résultats sur les fonctionnelles globales du Chapitre 3

Dans ce paragraphe, nous allons synthétiser les résultats du chapitre 3 qui correspond à l'article [54] paru dans *Electronic Journal of Probability*.

Le but de ce travail a été dans un premier temps, de prouver la conjecture posée par Fill et Janson [88] et donc de s'intéresser à l'étude asymptotique de fonctionnelles additives associées à des fonction de péage du type $b_n = n^\beta$ pour $n \in \mathbb{N}^*$ et $\beta > 0$ pour des arbres binaires sous le modèle de Catalan. Nous avons en fait montré un résultat plus fort permettant d'atteindre des fonctionnelles plus générales. La convergence des fonctionnelles associées aux fonctions de péage du type puissance sera un cas particulier d'application. Nous nous sommes ensuite intéressés à des résultats similaires mais pour les arbres simplement générés.

1.5.1 Résultats pour les arbres binaires sous le modèle de Catalan

Contexte

Soit $(T_n : n \in \mathbb{N}^*)$ une suite d'arbres binaires obtenus comme sous-arbres de l'arbre continu brownien selon la procédure d'échantillonnage décrite dans la partie 1.3.3. On rappelle que T_n a même loi qu'un arbre binaire à n nœuds internes sous le modèle de Catalan (i.e qu'un arbre choisi uniformément dans l'ensemble des arbres binaires complets et ordonnés ayant n nœuds internes et donc $n + 1$ feuilles). En particulier, T_n possède $2n + 1$ nœuds. Pour $v \in T_n$, on note $T_{n,v} = (T_n)_v$ le sous-arbre de T_n au dessus du nœud v .

Espaces fonctionnels

Soit I un intervalle de \mathbb{R} de mesure de Lebesgue strictement positive. On note $\mathcal{B}(I)$ l'espace des fonctions mesurables de I dans \mathbb{R} . On note $\mathcal{C}(I)$ l'espace des fonctions continues de I dans \mathbb{R} . Pour $f \in \mathcal{B}(I)$, on note $\|f\|_\infty$ la norme infinie et par $\|f\|_{\text{esssup}}$ la norme supremum essentiel de $|f|$ sur I . Les deux supremum coïncident quand f est continue.

Des mesures aléatoires

Pour $n \in \mathbb{N}^*$, on définit la mesure aléatoire pondérée A_n sur $[0, 1]$ définie pour tout $f \in \mathcal{B}([0, 1])$ par :

$$A_n(f) = |T_n|^{-3/2} \sum_{v \in T_n} |T_{n,v}| f\left(\frac{|T_{n,v}|}{|T_n|}\right).$$

Pour $h \in \mathcal{C}_+([0, 1])$, on définit la longueur de l'excursion de h au dessus du niveau $r \in \mathbb{R}_+$ et passant par $s \in [0, 1]$ par :

$$\sigma_{r,s}(h) = \int_0^1 dt \mathbf{1}_{\{m_h(s,t) \geq r\}},$$

où pour $s, t \in [0, 1]$, $m_h(s, t) = \inf_{u \in [s \wedge t, s \vee t]} h(u)$, voir figure 1.6.

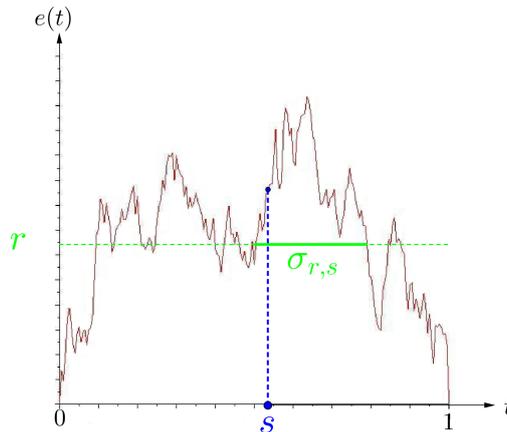


FIGURE 1.6 – Longueur d'une sous-excursion de e au dessus du niveau r et enjambant s .

Pour $h \in \mathcal{C}_+([0, 1])$, on considère aussi la mesure Φ_h sur $[0, 1]$ définie par :

$$\Phi_h(f) = \int_0^1 ds \int_0^{h(s)} dr f(\sigma_{r,s}(h)), \quad f \in \mathcal{B}([0, 1]).$$

On munit l'espace des mesures finies positives sur $[0, 1]$ de la topologie de la convergence étroite.

Une convergence presque sûre et ses fluctuations

On obtient la convergence presque sûre pour toutes les fonctions continues sur $[0, 1]$ qui explosent de manière contrôlée en 0 :

Théorème A. *On a presque sûrement, pour tout $f \in \mathcal{B}([0, 1])$, continue sur $(0, 1]$ et telle qu'il existe $a \in [0, 1/2)$ tel que $\lim_{x \rightarrow 0^+} x^a f(x) = 0$:*

$$A_n(f) \xrightarrow[n \rightarrow \infty]{} \sqrt{2\alpha} \Phi_e(f).$$

Remarque. La fonctionnelle de forme ne peut être atteinte par nos résultats : en effet, la fonction f définie par $f(x) = \log(x)/x$ ne satisfait pas les hypothèses du théorème A.

La preuve repose sur l'approximation de l'arbre brownien réel \mathcal{T} par un sous-arbre binaire $\mathcal{T}_{[n]}$, voir la construction rappelée dans la partie 1.3.3, puis sur l'approximation dans l'arbre marqué associé $\tilde{\mathcal{T}}_n$, des longueurs de branche $(h_{n,v} : v \in \mathcal{T}_n)$ par leur valeur moyenne que l'on déduit de (1.3). La mesure aléatoire A_n est ainsi approchée par plusieurs mesures aléatoires et chacune des approximations successives est contrôlée en norme L^2 .

On obtient également les fluctuations associées à cette convergence presque sûre :

Théorème B. *Soit $f \in \mathcal{C}([0, 1])$ une fonction localement lipschitzienne continue sur $(0, 1]$ telle qu'il existe $a \in (0, 1)$ tel que $\|x^a f'\|_{\text{esssup}}$ soit finie. On a la convergence en loi suivante :*

$$\left(|\mathcal{T}_n|^{1/4} (A_n - \sqrt{2\alpha} \Phi_e)(f), A_n \right) \xrightarrow[n \rightarrow \infty]{(d)} \left((2\alpha)^{1/4} \sqrt{\Phi_e(xf^2)} G, \sqrt{2\alpha} \Phi_e \right),$$

où G est une variable aléatoire de loi normale centrée, réduite indépendante de l'excursion brownienne e .

On constate dans la preuve que les fluctuations proviennent de l'estimation des longueurs de branche $(h_{n,v} : v \in \mathcal{T}_n)$ par leur moyenne dans l'arbre marqué $\tilde{\mathcal{T}}_n$ et non pas de l'approximation de l'arbre continu brownien \mathcal{T} par le sous-arbre $\mathcal{T}_{[n]}$.

Application aux fonctions de coût du type puissance

Dans toute cette partie, on considère le cas $\alpha = 2$ de sorte que e est l'excursion brownienne normalisée B_{ex} .

On applique les théorèmes A et B à la fonction f définie par $f(x) = x^{\beta-1}$, pour tout $x \in [0, 1]$, avec $\beta > 0$. On pose $Z_\beta := \Phi_{B_{\text{ex}}}(x^{\beta-1})$ et on définit une suite de variables aléatoires $(Z_\beta^{(n)} : n \in \mathbb{N}^*)$ par :

$$Z_\beta^{(n)} := A_n(x^{\beta-1}) = |\mathcal{T}_n|^{-(\beta+\frac{1}{2})} \sum_{v \in \mathcal{T}_n} |\mathcal{T}_{n,v}|^\beta, \quad \forall n \in \mathbb{N}^*,$$

D'après les hypothèses du théorème A, on obtient directement la convergence presque sûre pour tout $\beta > \frac{1}{2}$. Pour $0 < \beta \leq \frac{1}{2}$, un argument de convergence monotone en β est nécessaire. Nous obtenons finalement la convergence presque sûre suivante :

Proposition C. *On a presque sûrement pour tout $\beta > 0$:*

$$\lim_{n \rightarrow +\infty} Z_\beta^{(n)} = 2Z_\beta.$$

On montre également que pour $\beta > 1$,

$$\begin{aligned} 2Z_\beta &= \beta \int_0^1 \left[t^{\beta-1} + (1-t)^{\beta-1} \right] B_{\text{ex}}(t) dt \\ &\quad - \beta(\beta-1) \int_{0 < s < t < 1} (t-s)^{\beta-2} [B_{\text{ex}}(s) + B_{\text{ex}}(t) - 2m_{B_{\text{ex}}}(s,t)] ds dt, \end{aligned} \quad (1.6)$$

et que $2Z_1 = 2 \int_0^1 B_{\text{ex}}(s) ds$. Ces formules correspondent également à celles données par Fill et Janson [88]. Elle se simplifient dans le cas où $\beta > 1$ à :

$$2Z_\beta = \beta(\beta - 1) \int_{[0,1]^2} |t - s|^{\beta-2} m_{B_{\text{ex}}}(s, t) ds dt.$$

Fill et Janson [88] conjecturent que l'égalité (1.6) est également vraie pour tout $1/2 < \beta < 1$. Pour $\beta > 1/2$, Fill et Kapur [89] ont montré que $Z_\beta^{(n)}$ converge en loi vers $2Z_\beta$ et ont obtenu tous les moments de Z_β pour $\beta > 1/2$, voir proposition 3.5. Pour $0 < \beta \leq 1/2$, ils obtiennent également une convergence en loi mais avec une autre normalisation, voir propositions 3.5 et 3.9.

On montre également que Z_β et son moment d'ordre 1 sont presque sûrement finis si $\beta > 1/2$. Dans ce cas, on obtient l'expression du moment d'ordre 1, retrouvant ainsi l'expression donnée par Fill et Kapur [89]. Si $0 < \beta \leq 1/2$ alors Z_β est presque sûrement infinie. On met ainsi en évidence une transition de phase en $\beta = 1/2$.

La proposition C permet en particulier de retrouver, pour $\beta = 1$ et $\beta = 2$, les convergences classiques, voir par exemple Aldous [9] et Janson [113] :

$$|\mathbb{T}_n|^{-\frac{3}{2}} \sum_{v \in \mathbb{T}_n} |\mathbb{T}_{n,v}| \xrightarrow[n \rightarrow +\infty]{p.s.} 2 \int_0^1 B_{\text{ex}}(s) ds$$

et

$$|\mathbb{T}_n|^{-\frac{5}{2}} \sum_{v \in \mathbb{T}_n} |\mathbb{T}_{n,v}|^2 \xrightarrow[n \rightarrow +\infty]{p.s.} 4 \int_{0 \leq s \leq t \leq 1} m_{B_{\text{ex}}}(s, t) ds dt.$$

Nous terminons par les fluctuations associées à ce principe d'invariance :

Proposition D. *Pour tout $\beta \geq 1$, on a la convergence en loi suivante :*

$$\left(|\mathbb{T}_n|^{1/4} (Z_\beta^{(n)} - 2Z_\beta), Z_\beta^{(n)} \right) \xrightarrow[n \rightarrow +\infty]{(d)} \left(\sqrt{2Z_{2\beta}} G, 2Z_\beta \right),$$

où G est une variable aléatoire de loi normale centrée, réduite indépendante de l'excursion brownienne e .

1.5.2 Résultats pour les arbres simplement générés

Que se passe-t-il maintenant si l'arbre binaire à n nœuds internes sous le modèle de Catalan \mathbb{T}_n est remplacé par un arbre simplement généré $\tau^{(p)}$ à p nœuds ?

Contexte

Soit $\tau^{(p)}$ un arbre simplement généré à p nœuds de fonction poids \mathbf{p} . On suppose que \mathbf{p} est une probabilité critique sur \mathbb{N} , avec $0 < \mathbf{p}(0) \leq \mathbf{p}(1) + \mathbf{p}(0) < 1$ et qui appartient au domaine d'attraction d'une loi stable symétrique d'exposant de Laplace $\psi(\lambda) = \kappa \lambda^\gamma$ avec $\gamma \in (1, 2]$ et $\kappa > 0$, et de suite normalisée $(a_p, p \in \mathbb{N}^*)$. Pour $v \in \tau^{(p)}$, on note $\tau_v^{(p)}$ le sous-arbre de $\tau^{(p)}$ au dessus du nœud v .

Convergence d'une mesure aléatoire

De manière similaire à la mesure aléatoire pondérée A_n , on définit la mesure aléatoire finie non normalisée, pour tout $f \in \mathcal{B}([0, 1])$ par :

$$\mathcal{A}_{\tau^{(p)}}(f) = \sum_{v \in \tau^{(p)}} |\tau_v^{(p)}| f \left(\frac{|\tau_v^{(p)}|}{|\tau^{(p)}|} \right) = \sum_{v \in \tau^{(p)}} |\tau_v^{(p)}| f \left(\frac{|\tau_v^{(p)}|}{p} \right).$$

On rappelle que H est le processus hauteur associé à l'arbre de Lévy continu, voir paragraphe 1.2.3. On obtient la convergence en loi de la mesure $\mathcal{A}_{\tau^{(p)}}$ correctement normalisée.

Théorème E. *Soit \mathbf{p} une loi de probabilité critique sur \mathbb{N} , avec $0 < \mathbf{p}(0) \leq \mathbf{p}(1) + \mathbf{p}(0) < 1$, qui appartient au domaine d'attraction d'une loi stable symétrique d'exposant de Laplace $\psi(\lambda) = \kappa\lambda^\gamma$ avec $\gamma \in (1, 2]$ et $\kappa > 0$, et de suite normalisée $(a_p, p \in \mathbb{N}^*)$. Soit τ un arbre de Galton-Watson de loi de distribution \mathbf{p} , et $\tau^{(p)}$ de même loi que τ conditionnellement à $\{|\tau| = p\}$. On a la convergence en loi suivante :*

$$\frac{a_p}{p^2} \mathcal{A}_{\tau^{(p)}} \xrightarrow[p \rightarrow +\infty]{(d)} \Phi_H,$$

où on a muni l'espace des mesures positives de la topologie de la convergence étroite et où la convergence a lieu le long de toutes sous-suites infinies de p telles que $\mathbb{P}(|\tau| = p) > 0$.

En particulier, on a la convergence en loi suivante : pour tout $f \in \mathcal{C}([0, 1])$,

$$\frac{a_p}{p^2} \sum_{v \in \tau^{(p)}} |\tau_v^{(p)}| f\left(\frac{|\tau_v^{(p)}|}{p}\right) \xrightarrow[p \rightarrow +\infty]{(d)} \Phi_H(f).$$

La preuve repose sur la convergence globale des arbres de Galton-Watson renormalisés vers des arbres de Lévy en utilisant la convergence du processus de contour, voir la partie 1.2.3 concernant les travaux de Le Gall et Le Jan [136], Duquesne et Le Gall [70] et Duquesne [69].

Remarque. Si \mathbf{p} a une variance σ^2 finie alors on peut choisir $a_p = \sqrt{p}$ et H égale à $(2/\sigma)B_{\text{ex}}$.

Exemples d'applications :

Comme pour les arbres binaires sous le modèle de Catalan, on peut appliquer le théorème E à la fonction f définie par $f(x) = x^{\beta-1}$ pour $x \in [0, 1]$ et $\beta \geq 1$.

Sous les conditions du théorème E, il existe une suite $(a_p, p \in \mathbb{N}^*)$ tel que $\mathbb{P}(|\tau| = p) > 0$ telle que l'on a la convergence en loi suivante :

$$\frac{a_p}{p^{\beta+1}} \sum_{v \in \tau} |\tau_v^{(p)}|^\beta \xrightarrow[p \rightarrow +\infty]{(d)} Z_\beta^H = \Phi_H(x^{\beta-1}), \text{ pour tout } \beta \geq 1. \quad (1.7)$$

Maintenant, si \mathbf{p} a une variance σ^2 finie alors on peut choisir $a_p = \sqrt{p}$ et H égale à $(2/\sigma)B_{\text{ex}}$. En particulier, en utilisant le fait que $\Phi_{cB_{\text{ex}}} = c\Phi_{B_{\text{ex}}}$ où B_{ex} est l'excursion brownienne normalisée, on obtient les convergences pour les modèles suivants (voir tableau 1.1 pour les différents choix de σ^2) et pour tout $\beta \geq 1$:

- Arbres de Cayley à p nœuds :

$$\frac{1}{p^{\beta+\frac{1}{2}}} \sum_{v \in \mathbf{t}} |\tau_v^{(p)}|^\beta \xrightarrow[p \rightarrow +\infty]{(d)} 2\Phi_{B_{\text{ex}}}(x^{\beta-1}).$$

- Arbres de Catalan à p nœuds :

$$\frac{1}{p^{\beta+\frac{1}{2}}} \sum_{v \in \mathbf{t}} |\tau_v^{(p)}|^\beta \xrightarrow[p \rightarrow +\infty]{(d)} \sqrt{2}\Phi_{B_{\text{ex}}}(x^{\beta-1}).$$

On retrouve également la convergence des fonctionnelles associées aux fonctions péages de type puissance pour les arbres binaires sous le modèle de Catalan (la convergence est ici plus faible puisqu'elle est établie pour tout $\beta \geq 1$ tandis qu'elle l'était pour tout $\beta > 0$ précédemment).

1.5.3 Extensions possibles

Transition de phase pour les arbres simplement générés

Dans le cas des arbres binaires sous le modèle de Catalan, nous observons une transition de phase en $\beta = 1/2$ pour les fonctionnelles additives liées à des fonctions péages monômes $b_n = n^\beta$ avec $\beta > 0$. Quand $\beta > 1/2$, la fonctionnelle normalisée converge presque sûrement vers une limite Z_β finie tandis que quand $0 < \beta \leq 1/2$, cette limite est infinie. Ainsi pour $0 < \beta \leq 1/2$ la convergence n'est plus pertinente, voir Fill et Kapur [89], partie 3.4 pour la normalisation adaptée.

On aimerait montrer une transition de phase similaire pour les fonctionnelles additives pour des fonctions péages monômes $b_n = n^\beta$ avec $\beta > 0$ liées à des arbres simplement générés. On conjecture que sous les conditions du théorème E, la transition de phase est en $\beta = 1/\gamma$. En effet, on a montré la convergence en loi de ces fonctionnelles additives pour tout $\beta \geq 1$, vers une limite notée Z_β^H , voir (1.7). De plus, on montre que presque sûrement pour tout $0 < \beta \leq 1/\gamma$, Z_β^H est infinie tandis que presque sûrement, pour tout $\beta > 1/\gamma$, Z_β^H est finie, voir le lemme 3.17 du chapitre 3. Ainsi, pour prouver la transition de phase, il faudrait montrer que la convergence des fonctionnelles additives a également lieu pour $1/\gamma < \beta < 1$.

Fonctionnelles de coût dites asymétriques

Soit T_n un arbre binaire ayant n nœuds internes. On aimerait s'intéresser à des fonctionnelles dites asymétriques dont la fonction péage dépendrait non plus seulement du cardinal de l'arbre entier mais du cardinal de ses sous-arbres de gauche et de droite.

Un exemple de telles fonctionnelles est donné par l'indice de déséquilibre (« imbalance parameter » en anglais) qui compte le nombre de nœuds qui ont leurs sous-arbres de gauche et de droite de même taille, voir Devroye [60] :

$$\sum_{v \in T_n} \mathbf{1}_{\{|L_{n,v}|=|R_{n,v}|\}},$$

où $L_{n,v}$ est le sous-arbre de gauche de T_n au dessus du nœud v et $R_{n,v}$ est le sous-arbre de droite de T_n au dessus du nœud v .

Il serait donc naturel de s'intéresser à des fonctionnelles qui mesurent l'asymétrie des arbres. Pour $n \in \mathbb{N}^*$, on peut ainsi définir la mesure aléatoire pondérée B_n sur $[0, 1]^2$ définie pour tout $f \in \mathcal{B}([0, 1]^2)$ par :

$$B_n(f) = |T_n|^{-2} \sum_{v \in T_n} |L_{n,v}| |R_{n,v}| f \left(\frac{|L_{n,v}|}{|T_n|}, \frac{|R_{n,v}|}{|T_n|} \right),$$

On conjecture la convergence de cette mesure aléatoire pour les arbres binaires sous le modèle de Catalan.