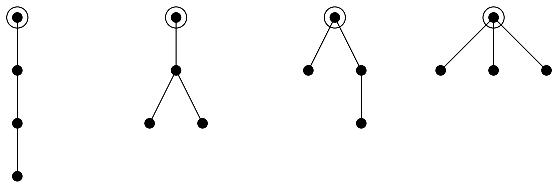
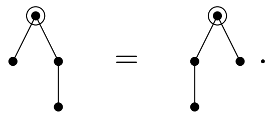
Mesure de Plancherel sur les arbres enracinés

mots-clés : mesure stationnaire, arbres enracinés, relation de commutation.

L'objectif de ce texte est d'étudier une mesure de probabilité et une marche aléatoire sur l'ensemble $\mathfrak{T}(n)$ des *arbres enracinés* de taille N. Par arbre enraciné de taille N, on entend un graphe connexe non vide $G = (V, E)$ avec $N = |V|$ sommets, $N - 1 = |E|$ arêtes non orientées, et un sommet $v \in V$ distingué qui est la *racine*. Deux arbres enracinés $G_1 = (v_1, V_1, E_1)$ et $G_2 = (v_2, V_2, E_2)$ sont identifiés s'il existe une bijection $\psi : V_1 \to V_2$ telle que $\psi(v_1) = v_2$ et telle que $\{a, b\} \in E_1$ si et seulement si $\{\psi(a), \psi(b)\} \in E_2$. Par exemple, l'ensemble $\mathfrak{T}(4)$ des arbres enracinés de taille 4 contient 4 arbres :



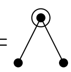
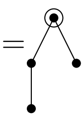
Remarquons en particulier qu'avec notre définition d'arbre enraciné,



Dans la suite, les arbres enracinés seront dessinés avec la racine tout en haut, et les *feuilles* d'un arbre sont bien sûr les sommets de valence 1, qui apparaissent en bas de l'arbre. On note $\bullet$ l'unique arbre enraciné de taille $N = 1$. On rappelle que si w, z sont deux sommets voisins d'un arbre enraciné en v, on dit que z est un *fils* de w si et seulement si, pour la distance de graphe, $\mathrm{dist}(z, v) = \mathrm{dist}(w, v) + 1$.

1 Greffes et élagages

Si $t \in \mathfrak{T}(N)$ et $t' \in \mathfrak{T}(N+1)$, on note $t \nearrow t'$ si l'on peut obtenir t à partir de t' en retirant une feuille à t' (élagage). On note alors $d(t, t')$ le *nombre d'élagage*, qui est le nombre de feuilles de t' qui après élagage donnent t. Ce nombre peut être différent du *nombre de greffe* $u(t, t')$, qui est le nombre de sommets de t auxquels on peut greffer une feuille pour obtenir t'. Par exemple, avec les arbres

$t =$  $t' =$ 

on a $d(t, t') = 1$ et $u(t, t') = 2$. Un lien entre les deux quantités peut être établi en introduisant le *groupe de symétrie* d'un arbre $t = (v, V, E)$, qui est le groupe $\mathrm{SG}(t)$ des permutations $\sigma : V \to V$ telles que $\sigma(v) = v$, et qui respectent les relations de descendance : w est un fils de v si et seulement si $\sigma(w)$ est un fils de $\sigma(v)$. Le *facteur de symétrie* de t est le cardinal $|\mathrm{SG}(t)|$ du groupe de symétrie. Une application de la formule des orbites permet de montrer que si $t \nearrow t'$, alors on a

$$d(t, t')\, |\mathrm{SG}(t)| = u(t, t')\, |\mathrm{SG}(t')|. \tag{1}$$

Dans l'exemple précédent, $\mathrm{SG}(t)$ contient deux éléments (on peut échanger les deux sous-arbres issus des fils de la racine), tandis que $\mathrm{SG}(t')$ contient seulement l'identité.

Plus généralement, considérons deux arbres $t \in \mathfrak{T}(N)$ et $t' \in \mathfrak{T}(N+P)$ avec $N, P \geq 0$. On note $t \leq t'$ s'il existe au moins une suite d'arbres $t_1, t_2, \ldots, t_{P-1}$ telle que

$$t \nearrow t_1 \nearrow t_2 \nearrow \cdots \nearrow t_{P-1} \nearrow t'.$$

Les définitions des nombres de greffe et d'élagage s'étendent à ce cadre de la façon suivante : on pose

$$d(t,t') = \sum_{t \nearrow t_1 \nearrow t_2 \nearrow \cdots \nearrow t_{P-1} \nearrow t'} d(t,t_1)\, d(t_1,t_2)\, \cdots\, d(t_{P-1},t');$$
$$u(t,t') = \sum_{t \nearrow t_1 \nearrow t_2 \nearrow \cdots \nearrow t_{P-1} \nearrow t'} u(t,t_1)\, u(t_1,t_2)\, \cdots\, u(t_{P-1},t'),$$

étant entendu que les sommes sont nulles si l'on n'a pas $t \leq t'$. Le nombre d'élagage $d(t,t')$ (respectivement, le nombre de greffe $u(t,t')$) est donc le nombre de façons possibles d'obtenir t à partir de t' en retirant successivement des feuilles (respectivement, en ajoutant successivement des feuilles). La relation (1) énoncée pour les paires d'arbres (t,t') avec $t \nearrow t'$ s'étend par récurrence sur $P = |t'| - |t|$ à toutes les paires d'arbres.

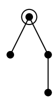
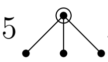
2 Mesure de Plancherel

Pour tout arbre enraciné t, on pose $d(t) = d(\bullet, t)$ et $u(t) = u(\bullet, t)$. Soit $N \geq 1$. La mesure de Plancherel sur $\mathfrak{T}(N)$ est la mesure

$$\mathbb{P}_N[t] = \frac{d(t)\, u(t)}{\prod_{i=2}^{N} \binom{i}{2}} = \frac{(d(t))^2}{|\mathrm{SG}(t)|\, \prod_{i=2}^{N} \binom{i}{2}}. \tag{2}$$

Théorème 1. *Pour tout $N \geq 1$, $\mathbb{P}_N$ est une mesure de probabilité sur $\mathfrak{T}(N)$ appelée* mesure de Plancherel.

Introduisons quelques notations. Étant donné un ensemble fini ou dénombrable $S = \{s_i\}_{i \in I}$, on note $\mathbb{R}[S]$ l'ensemble des combinaisons linéaires formelles d'éléments de S ; en particulier, $\mathbb{R}[\mathfrak{T}(N)]$ est un espace vectoriel de dimension $|\mathfrak{T}(N)|$ et dont une base est donnée par les arbres enracinés de taille N. Un exemple d'élément de $\mathbb{R}[\mathfrak{T}(4)]$ est

$x = 3$  $- 5$  .

Introduisons les trois applications linéaires $\mathscr{R}_N = N\, \mathrm{id}_{\mathbb{C}[\mathfrak{T}(N)]}$,

$$\mathscr{D}_N : \mathbb{R}[\mathfrak{T}(N+1)] \to \mathbb{R}[\mathfrak{T}(N)] \qquad ; \qquad \mathscr{U}_N : \mathbb{R}[\mathfrak{T}(N)] \to \mathbb{R}[\mathfrak{T}(N+1)]$$
$$t' \mapsto \sum_{t \nearrow t'} d(t,t')\, t \qquad\qquad\qquad t \mapsto \sum_{t \nearrow t'} u(t,t')\, t'.$$

On convient que $\mathscr{D}_0 = \mathscr{U}_0 = 0$. Alors, on a pour tout $N \geq 1$ l'identité remarquable

$$\mathscr{D}_N \circ \mathscr{U}_N - \mathscr{U}_{N-1} \circ \mathscr{D}_{N-1} = \mathscr{R}_N. \tag{3}$$

Par ailleurs, si chaque espace $\mathbb{R}[\mathfrak{T}(N)]$ est muni du produit scalaire $\langle t \,|\, t' \rangle = |\mathrm{SG}(t)|\, 1_{(t=t')}$, alors l'identité (1) implique que les opérateurs $\mathscr{D}_N$ et $\mathscr{U}_N$ sont adjoints pour tout $N \geq 1$.

Soit $\mathfrak{T} = \bigsqcup_{N\geq 1} \mathfrak{T}(N)$, et $\mathbb{R}[\mathfrak{T}] = \bigoplus_{n\geq 1} \mathbb{R}[\mathfrak{T}(N)]$ l'ensemble des sommes formelles d'arbres enracinés de tailles arbitraires. On étend à $\mathbb{R}[\mathfrak{T}]$ le produit scalaire précédemment défini sur chaque $\mathbb{R}[\mathfrak{T}(N)]$, en décrétant que les sous-espaces $\mathbb{R}[\mathfrak{T}(N)]$ sont deux-à-deux orthogonaux. Les applications linéaires $\mathscr{D}, \mathscr{U}, \mathscr{R} : \mathbb{R}[\mathfrak{T}] \to \mathbb{R}[\mathfrak{T}]$ sont définies par les conditions suivantes :

$$\mathscr{D}_{|\mathbb{R}[\mathfrak{T}(N+1)]} = \mathscr{D}_N \qquad ; \qquad \mathscr{U}_{|\mathbb{R}[\mathfrak{T}(N)]} = \mathscr{U}_N \qquad ; \qquad \mathscr{R}_{|\mathbb{R}[\mathfrak{T}(N)]} = \mathscr{R}_N,$$

et l'identité remarquable (3) donne la relation de commutation

$$\mathscr{D} \circ \mathscr{U} - \mathscr{U} \circ \mathscr{D} = \mathscr{R}.$$

En utilisant récursivement cette relation de commutation, on voit facilement que pour tout $P \geq 1$ et tout mot w en les opérateurs $\mathscr{U}$ et $\mathscr{D}$ (par exemple, $w = \mathscr{D}^3\mathscr{U}^2\mathscr{D}\mathscr{U}$), il existe des polynômes $(c_{w,a,b}(P))_{a,b\geq 0}$ tels que

$$w_{|\mathbb{R}[\mathfrak{T}(P)]} = \sum_{a,b\geq 0} c_{w,a,b}(P)\, (\mathscr{U}^a \circ \mathscr{D}^b)_{|\mathbb{R}[\mathfrak{T}(P)]}. \tag{4}$$

Par exemple,

$$\begin{aligned}(\mathscr{D}\mathscr{U})_{\mathbb{R}[\mathfrak{T}(P)]} &= (\mathscr{U}\mathscr{D})_{\mathbb{R}[\mathfrak{T}(P)]} + P\,\mathrm{id}_{\mathbb{R}[\mathfrak{T}(P)]};\\ (\mathscr{D}^2\mathscr{U}^2)_{\mathbb{R}[\mathfrak{T}(P)]} &= (\mathscr{U}^2\mathscr{D}^2)_{\mathbb{R}[\mathfrak{T}(P)]} + (4P-1)(\mathscr{U}\mathscr{D})_{\mathbb{R}[\mathfrak{T}(P)]} + 2P^2\,\mathrm{id}_{\mathbb{R}[\mathfrak{T}(P)]}.\end{aligned}$$

On admet dans ce qui suit que la décomposition (4) est unique. Ceci implique facilement :

Lemme 2. *Les polynômes $c_{w,a,b}(P)$ vérifient la relation de récurrence :*

$$c_{\mathscr{D}w,a,b}(P) = c_{w,a,b-1}(P) + ((P-b) + (P-b+1) + \cdots + (P-b+a))\, c_{w,a+1,b}(P).$$

Preuve du théorème 1. On va calculer $S_N = \sum_{t\in\mathfrak{T}(N)} d(t)\, u(t)$. La discussion qui précède permet de réécrire cette quantité comme suit :

$$\begin{aligned}S_N &= \sum_{t\in\mathfrak{T}(N)} d(\bullet,t)\, u(\bullet,t) = \sum_{t\in\mathfrak{T}(N)} \left\langle \bullet \,\middle|\, \mathscr{D}^{N-1}(t)\right\rangle u(\bullet,t) = \sum_{t\in\mathfrak{T}(N)} u(\bullet,t) \left\langle \mathscr{U}^{N-1}(\bullet) \,\middle|\, t\right\rangle\\ &= \left\langle \mathscr{U}^{N-1}(\bullet) \,\middle|\, \mathscr{U}^{N-1}(\bullet)\right\rangle = \left\langle \mathscr{D}^{N-1}\mathscr{U}^{N-1}(\bullet) \,\middle|\, \bullet\right\rangle\\ &= \sum_{a,b\geq 0} c_{\mathscr{D}^{N-1}\mathscr{U}^{N-1},a,b}(1) \left\langle \mathscr{U}^a\mathscr{D}^b(\bullet) \,\middle|\, \bullet\right\rangle\\ &= c_{\mathscr{D}^{N-1}\mathscr{U}^{N-1},0,0}(1),\end{aligned}$$

la dernière identité provenant du fait que $\mathscr{D}(\bullet) = 0$. La relation de récurrence du lemme 2 donne

$$\begin{aligned}c_{\mathscr{D}^{N-1}\mathscr{U}^{N-1},0,0}(1) &= \binom{N}{2} c_{\mathscr{D}^{N-2}\mathscr{U}^{N-1},1,0}(1) = \binom{N}{2}\binom{N-1}{2} c_{\mathscr{D}^{N-3}\mathscr{U}^{N-1},2,0}(1) = \cdots\\ &= \left(\prod_{i=2}^{N}\binom{i}{2}\right) c_{\mathscr{U}^{N-1},N-1,0}(1) = \prod_{i=2}^{N}\binom{i}{2}.\end{aligned}$$

□

3 La chaîne de Markov sur les arbres

Fixons un entier $N \geq 2$. On définit des probabilités de transitions $Q_{N,d} : \mathfrak{T}(N)\times\mathfrak{T}(N-1) \to [0,1]$ et $Q_{N,u} : \mathfrak{T}(N-1)\times\mathfrak{T}(N) \to [0,1]$ par les formules suivantes :

$$Q_{N,d}(t',t) = \frac{d(t)\, d(t,t')}{d(t')} \qquad ; \qquad Q_{N,u}(t,t') = \frac{d(t,t')\, u(t')}{\binom{N}{2}\, u(t)}.$$

Lemme 3. *Les matrices rectangulaires $Q_{N,d}$ et $Q_{N,u}$ sont stochastiques :*

$$\forall t' \in \mathfrak{T}(N), \quad \sum_{t\in\mathfrak{T}(N-1)} Q_{N,d}(t',t) = 1 \qquad ; \qquad \forall t \in \mathfrak{T}(N-1), \quad \sum_{t'\in\mathfrak{T}(N)} Q_{N,u}(t,t') = 1.$$

Démonstration. C'est évident pour la matrice $Q_{N,d}$. Pour la matrice $Q_{N,u}$, on va calculer pour tout arbre $t \in \mathfrak{T}(N-1)$ la somme

$$\begin{aligned}
S(t) &= \sum_{t' \,|\, t\nearrow t'} d(t,t')\, u(t') = \frac{1}{|\mathrm{SG}(t)|} \sum_{t'\in\mathfrak{T}(N)} u(t') \left\langle t \,\middle|\, \sum_{s\in\mathfrak{T}(N-1)} d(s,t')\, s \right\rangle \\
&= \frac{1}{|\mathrm{SG}(t)|} \sum_{t'\in\mathfrak{T}(N)} u(t')\, \langle t \mid \mathscr{D}(t')\rangle = \frac{1}{|\mathrm{SG}(t)|} \sum_{t'\in\mathfrak{T}(N)} u(t')\, \langle \mathscr{U}(t) \mid t'\rangle \\
&= \frac{1}{|\mathrm{SG}(t)|} \left\langle \mathscr{U}(t) \,\middle|\, \mathscr{U}^{N-1}(\bullet)\right\rangle = \frac{1}{|\mathrm{SG}(t)|} \left\langle t \,\middle|\, \mathscr{D}\mathscr{U}^{N-1}(\bullet)\right\rangle.
\end{aligned}$$

En utilisant comme précédemment la relation de commutation entre $\mathscr{D}$ et $\mathscr{U}$, on obtient :

$$\mathscr{D}\mathscr{U}^{N-1}(\bullet) = \binom{N}{2}\, \mathscr{U}^{N-2}(\bullet). \tag{5}$$

On en déduit que $S(t) = \binom{N}{2}\, u(t)$. □

La propriété de matrice stochastique reste valable pour le produit matriciel $Q_N = Q_{N,d}\, Q_{N,u}$, dont les coefficients sont :

$$Q_N(r,t) = \sum_{s\in\mathfrak{T}(N-1)} Q_{N,d}(r,s)\, Q_{N,u}(s,t).$$

Théorème 4. *Les mesures de Plancherel vérifient les relations :*

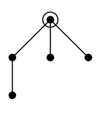
$$\mathbb{P}_N\, Q_{N,d} = \mathbb{P}_{N-1} \qquad ; \qquad \mathbb{P}_{N-1}\, Q_{N,u} = \mathbb{P}_N.$$

La chaîne de Markov sur $\mathfrak{T}(N)$ de matrice de transition Q_N est irréductible, apériodique et de mesure invariante $\mathbb{P}_N$. Cette mesure est réversible.

Ainsi, si l'on part d'un arbre enraciné $t \in \mathfrak{T}(N)$ et qu'on effectue un grand nombre de fois l'opération (élaguer une branche suivant la loi de transition $Q_{N,d}$, puis greffer une feuille suivant la loi de transition $Q_{N,u}$), alors on obtiendra un arbre enraciné aléatoire de loi proche de $\mathbb{P}_N$.

Questions

1. Un arbre enraciné peut être défini récursivement comme suit : c'est soit l'ensemble vide $\{\}$ (arbre trivial $\bullet$), soit un multi-ensemble fini d'arbres enracinés (l'ensemble des sous-arbres issus des fils de la racine, comptés avec multiplicités). Par exemple,

 $= \{\{\{\}\}, \{\}, \{\}\}.$

En utilisant cette description, montrer que les cardinaux $T_N = |\mathfrak{T}(N)|$ vérifient l'identité suivante :

$$\sum_{N=1}^{\infty} T_N\, x^N = x \prod_{N\geq 1} \frac{1}{(1-x^N)^{T_N}}$$

cette identité étant valable dans l'ensemble des séries formelles en la variable x. On pourra commencer par rappeler le développement en série de $\frac{1}{(1-z)^t}$ avec $t \in \mathbb{N}^*$.

2. Déduire des calculs de la question précédente un programme récursif `NombreArbres(N)` qui calcule T_N. Donner les valeurs de T_N pour $1 \leq N \leq 10$.

3. Soient t et t' deux arbres enracinés de tailles respectives N et $N+1$. On choisit une identification de $V(t)$ avec un sous-ensemble de $V(t')$, et on note w, z les deux éléments de $V(t')$ tels que $V(t') = \{z\} \sqcup V(t)$, et tels que le père de z dans t' est $w \in V(t)$. Utiliser la formule des orbites pour montrer que

$$u(t,t') = \frac{|\mathrm{SG}(t)|}{|\mathrm{Fix}(w,t)|} \qquad ; \qquad d(t,t') = \frac{|\mathrm{SG}(t')|}{|\mathrm{Fix}(z,t')|},$$

où $\mathrm{Fix}(w,t) \subset \mathrm{SG}(t)$ est le sous-groupe de permutations fixant w, et $\mathrm{Fix}(z,t') \subset \mathrm{SG}(t')$ est le sous-groupe de permutations fixant z. Identifier les deux groupes $\mathrm{Fix}(w,t)$ et $\mathrm{Fix}(z,t')$, et en déduire la relation (1).

4. Expliquer pourquoi les deux formules dans l'équation (2) sont équivalentes. Calculer les probabilités sous la mesure de Plancherel de tous les arbres enracinés de taille 4.

5. Montrer que $\mathscr{D}_N$ et $\mathscr{U}_N$ sont adjoints pour les produits scalaires sur $\mathbb{R}[\mathfrak{T}(N)]$ et $\mathbb{R}[\mathfrak{T}(N+1)]$.

6. Établir l'identité remarquable (3). On pourra numéroter $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_N\}$ les sommets d'un arbre enraciné $t \in \mathfrak{T}(N)$, de sorte que les $v_{i>r}$ sont les feuilles de t. Notons $t_{\downarrow j>r}$ l'arbre obtenu à partir de t en supprimant la feuille v_j, et $t^{\uparrow i\geq 1}$ l'arbre obtenu à partir de t en rajoutant une feuille fille du sommet v_i. Exprimer en fonction des arbres $t_{\downarrow j}$ et $t^{\uparrow i}$ les deux expressions $\mathscr{D}_{N-1}(t)$ et $\mathscr{U}_N(t)$, et en déduire l'identité.

7. En utilisant l'unicité de la décomposition (4) et la relation de commutation $\mathscr{D}\mathscr{U} - \mathscr{U}\mathscr{D} = \mathscr{R}$, démontrer le lemme 2. On pourra commencer par étudier les opérateurs $(\mathscr{D}\mathscr{U}^a)_{\mathbb{R}[\mathfrak{T}(P)]}$ pour $a, P \geq 1$.

8. Expliquer pourquoi la matrice $Q_{N,d}$ est stochastique. Démontrer l'identité (5) à partir du lemme 2, et expliquer pourquoi elle implique que $S(t) = \binom{N}{2}\, u(t)$.

9. Démontrer entièrement le théorème 4. Peut-on utiliser ce théorème pour construire un arbre suivant *exactement* la mesure de Plancherel $\mathbb{P}_N$?

10. Proposer des programmes `Python` qui illustrent certains des résultats du texte. Si l'on veut programmer des arbres (ce n'est pas obligatoire), on pourra utiliser dans `Python/SageMath` la classe `RootedTree` (voir `http://doc.sagemath.org/html/en/reference/combinat/sage/combinat/rooted_tree.html`). On pourra aussi utiliser sans démonstration la formule suivante : pour $t \in \mathfrak{T}(N)$,

$$d(t) = \frac{N!}{\prod_{w\in V(t)} |t_w|}, \quad \text{où } t_w \text{ désigne le sous-arbre de } t \text{ issu du sommet } w.$$