



## ... .. l'arbre continu brownien

Nous présentons *l'arbre continu brownien*. Il s'agit d'un arbre continu fractal aléatoire, introduit par David Aldous au début des années 1990, qui est depuis devenu un objet important en probabilités et qui apparaît dans des contextes variés. Chemin faisant, ce sera l'occasion de présenter de jolis résultats mathématiques accessibles dès les premières années postbac.

- N. CURIEN
- I. KORTCHEMSKI

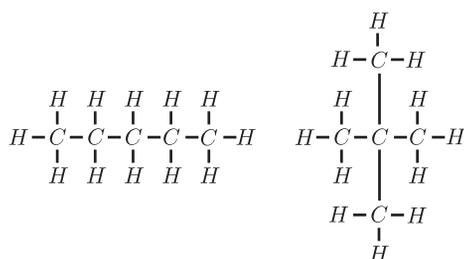
### 1. Introduction

À quoi une structure arborescente typique ressemble-t-elle ?

#### 1.1 – Souvenirs du lycée...

Prenons un exemple concret issu de la chimie. Considérons  $n$  atomes de carbone et  $2n + 2$  atomes d'hydrogène. Il est alors possible d'assembler ces atomes pour former une molécule sans cycle (un alcane) en créant 4 liaisons par atome de carbone et 1 liaison par atome d'hydrogène. Bien que toutes les molécules obtenues aient la même formule chimique  $C_nH_{2n+2}$ , leurs structures géométriques peuvent être bien différentes comme suggéré par la figure 1 (on dit que ce sont des isomères).

FIGURE 1 – Deux isomères du pentane



Combien vaut  $\text{Alc}_n$ , le nombre d'isomères d'un alcane à  $n$  atomes de carbone ? Il n'existe malheureusement pas de formule simple donnant l'expression de  $\text{Alc}_n$  en fonction de  $n$  mais on peut calcu-

ler les premières valeurs 1, 1, 1, 2, 3, 5, 9, 18... pour  $n = 1, 2, 3, 4, 5, 6, 7, 8...$

La question qui va nous intéresser ici est de comprendre la structure géométrique d'un alcane à  $n$  atomes de carbone *typique*. Plus précisément, nous pouvons considérer un isomère choisi uniformément au hasard parmi les  $\text{Alc}_n$  possibilités, et se demander quelle est sa géométrie (par exemple : quel est son diamètre vu comme graphe) ? Bien que les structures géométriques de ces isomères peuvent être très différentes (on peut imaginer un isomère très longiligne, ou au contraire un isomère ressemblant à une sorte d'étoile), nous verrons qu'en pratique ces comportements atypiques auront très peu de chance d'arriver.

#### 1.2 – Structures discrètes aléatoires

La question précédente s'inscrit dans un cadre plus général. Pour tout  $n \geq 1$ , imaginons un ensemble fini  $\mathbb{X}_n$  représentant un ensemble de structures discrètes de taille  $n$  ; si  $X_n$  est un élément de  $\mathbb{X}_n$  choisi uniformément au hasard, que peut-on dire de  $X_n$  lorsque  $n \rightarrow \infty$  ?

Pour répondre à cette question, il est utile d'avoir une manière simple de calculer  $\text{Card}(\mathbb{X}_n)$  ainsi qu'une méthode de simulation de  $X_n$ .

Nous allons esquisser dans la Section 2 des réponses à ces questions pour deux types de structures arborescentes : les arbres étiquetés et les arbres plans. Nous verrons que les algorithmes de simulation de ces arbres aléatoires sont aussi des outils bien pratiques pour leur étude mathé-

matique. Ces deux modèles sont en réalité très liés car asymptotiquement lorsque leur taille grandit, il n'est presque plus possible de distinguer un arbre plan aléatoire d'un arbre étiqueté aléatoire (ni même d'un alcane aléatoire) : ils convergent vers la même limite, qui est l'arbre continu brownien (c'est l'objet de la Section 3).

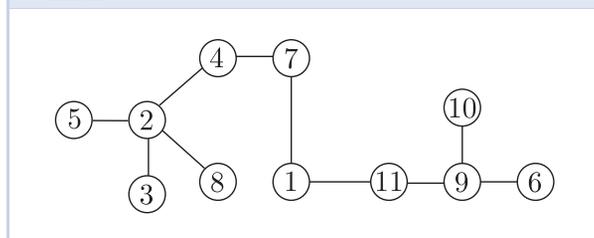
## 2. Cayley & Catalan

Dans cette partie, nous nous intéressons aux propriétés des arbres étiquetés (dits de Cayley) et des arbres plans (dits de Catalan).

### 2.1 – Les arbres de Cayley

Un arbre de Cayley à  $n$  sommets est, par définition, un arbre au sens de la théorie des graphes, c'est-à-dire un graphe sans cycle, et dont les sommets sont étiquetés de 1 à  $n$ . Voir la figure 2.

FIGURE 2 – Un arbre de Cayley sur 11 sommets



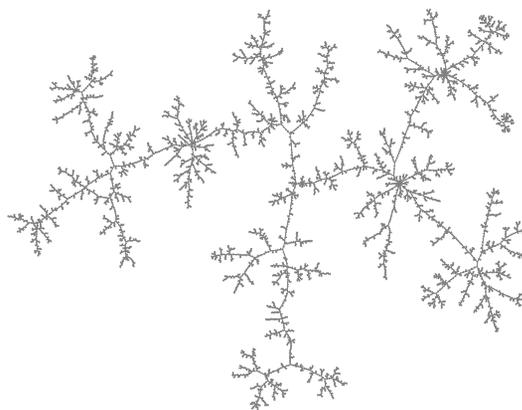
**Combinatoire.** Si on note  $Cay_n$  le nombre d'arbres de Cayley à  $n$  sommets, alors la suite  $(Cay_n)$  commence par les valeurs suivantes : 1, 1, 3, 16, 75, etc. La *formule de Cayley* (en fait due à Sylvester) fournit une expression simple de  $Cay_n$  en fonction de  $n$  :

$$Cay_n = n^{n-2}, \quad n \geq 1.$$

Il existe plusieurs preuves élégantes de ce joli résultat (voir par exemple sur Wikipedia).

**Simulation.** En suivant le plan de route esquissé dans la Section 1.2, intéressons-nous maintenant à la géométrie d'un arbre de Cayley  $\mathcal{T}_n$  tiré uniformément au hasard parmi les  $Cay_n$  arbres de Cayley à  $n$  sommets. Voici une simulation d'un arbre de Cayley

aléatoire à 10000 sommets représenté dans le plan obtenu avec le logiciel Mathematica en utilisant la fonction Graphplot :



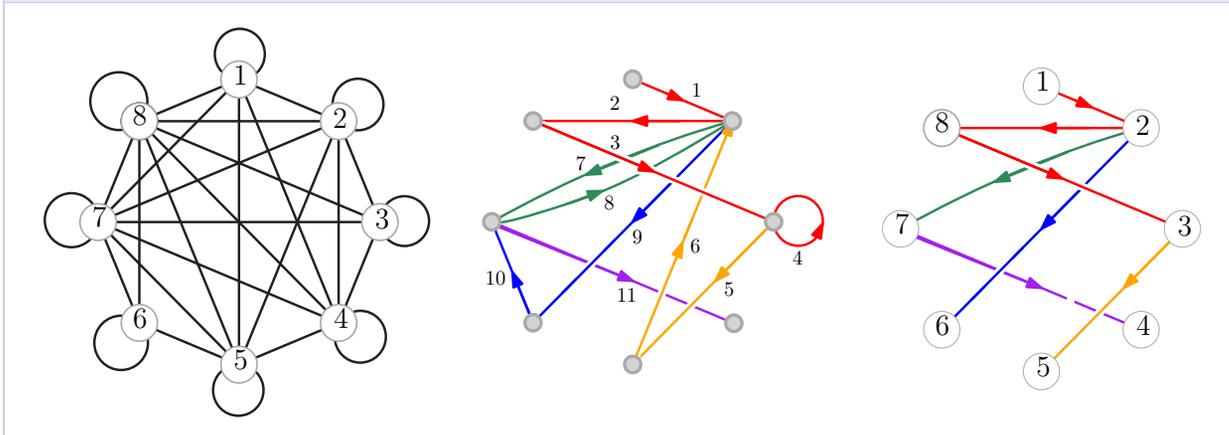
Comment une telle simulation a-t-elle été réalisée ? Avons-nous fait la liste des  $10000^{9998}$  arbres de Cayley à 10000 sommets avant d'en choisir un uniformément parmi eux ? Non, bien sûr. Il est possible de tirer un arbre de Cayley uniformément au hasard sans avoir à les énumérer grâce à l'algorithme d'Aldous & Broder découvert autour de 1990. Voici comment il fonctionne : on considère une marche aléatoire sur le graphe complet à  $n$  sommets  $\{1, 2, \dots, n\}$ , c'est-à-dire ni plus ni moins qu'une suite  $X_0, X_1, \dots$  de nombres indépendants et uniformes sur  $\{1, 2, 3, \dots, n\}$ . On construit alors itérativement un sous-arbre croissant :

- à  $n = 0$  le graphe est formé du seul sommet  $X_0$ ;
- pour  $n > 0$  si l'on visite un nouveau sommet à l'étape  $n$  alors on ajoute l'arête  $X_n \leftrightarrow X_{n-1}$  au sous-arbre.

Il est facile de voir qu'à chaque instant le graphe formé par l'ajout de ces arêtes reste un arbre et que le processus se stabilise quand la marche aléatoire a visité tous les sommets du graphe. Le *théorème d'Aldous-Broder*<sup>1</sup> dit que l'arbre aléatoire ainsi créé est en fait un arbre uniformément distribué sur  $Cay_n$ . Dans notre cas, le temps nécessaire pour visiter tous les sommets du graphe est le temps qu'il faut pour « collecter » les  $n$  nombres  $1, 2, \dots, n$  en tirant des variables uniformes  $X_0, X_1, \dots$  : c'est le problème du collecteur de coupons.

1. Plus généralement, cet algorithme permet de construire un arbre couvrant uniforme d'un graphe (fini connexe) donné.

FIGURE 3 – Création de l'arbre de Cayley à partir de la suite 1,2,8,3,3,5,2,7,2,6,7,4... sur le graphe complet à 8 sommets. La marche change de couleur dès qu'elle retombe sur un sommet déjà visité



**Application.** Maintenant que nous avons un procédé algorithmique efficace pour générer un arbre de Cayley uniforme à  $n$  sommets, intéressons-nous à une quantité géométrique simple : la longueur  $H_n$  de la première branche (en rouge sur la figure 3). Il est possible de démontrer que cette longueur peut être interprétée comme la longueur entre la racine et un sommet choisi uniformément au hasard dans l'arbre de Cayley.

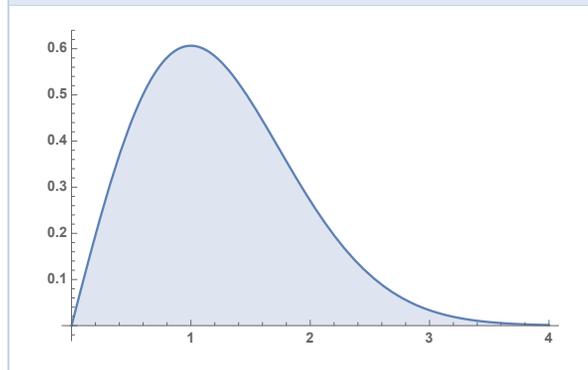
Cette branche grandit tant que le processus  $(X_n)$  découvre de nouveaux sommets et elle s'arrête dès que  $(X_n)$  tombe sur une valeur déjà vue. Ce problème est souvent posé lorsque  $n = 365$  : dans une classe d'école, cochez successivement les dates d'anniversaires d'élèves pris au hasard et arrêtez-vous dès qu'une date est cochée deux fois. Le paradoxe des anniversaires stipule qu'il suffit d'environ  $40 \approx 2 \times \sqrt{365}$  tirages avant de tomber deux fois sur un même anniversaire avec une probabilité d'environ 90% (c'est bien moins que la borne déterministe de  $365+1$ ). Faisons le calcul explicite dans notre cas. Dire que la première branche a au moins  $k$  sommets revient à dire que le deuxième sommet est différent du premier, ce qui arrive avec probabilité  $1 - 1/n$ , que le troisième est différent des deux premiers, probabilité  $1 - 2/n$ , et ainsi de suite. Par indépendance, on en déduit que pour tout  $k \geq 1$ ,

$$\mathbb{P}(H_n \geq k) = \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) \approx e^{-\frac{k^2}{2n}}.$$

Ainsi, tant que  $k \ll \sqrt{n}$  la probabilité ci-dessous est d'ordre 1 et devient négligeable quand  $k \gg \sqrt{n}$ . Quand  $k/\sqrt{n} \rightarrow x$ , cette probabilité converge

vers  $\exp(-x^2/2)$ . En termes probabilistes, on dit que lorsque  $n \rightarrow \infty$ , la variable aléatoire  $H_n/\sqrt{n}$  converge en loi vers une variable aléatoire réelle à densité, de densité  $x \mapsto x \exp(-x^2/2)$ . La loi de cette dernière variable aléatoire est appelée *loi de Rayleigh*.

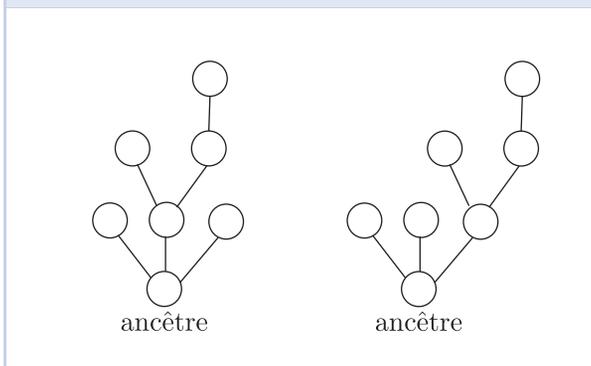
FIGURE 4 – Densité de la loi de Rayleigh, découverte par Lord Rayleigh lors de ses études sur la transmission du son dans les matériaux hétérogènes



## 2.2 – Les arbres de Catalan

Dans cette section nous allons nous intéresser aux arbres plans. Il s'agit d'arbres qui peuvent être utilisés pour modéliser la généalogie d'une famille (asexuée) partant d'un ancêtre et dont les enfants de chaque sommet sont ordonnés de gauche à droite (voir la figure 5 pour un exemple).

FIGURE 5 – Deux arbres plans différents (mais ayant la même structure de graphe)



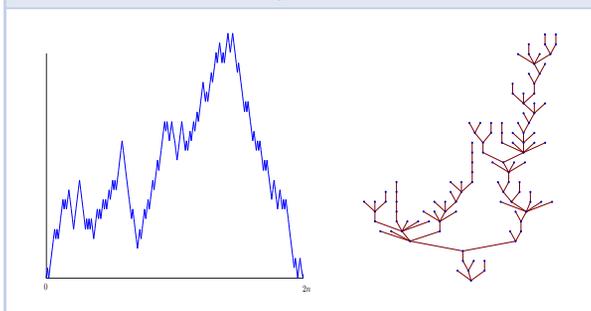
**Combinatoire.** Si on note  $Cat_n$  le nombre d'arbres plans à  $n + 1$  sommets, alors la suite  $(Cat_n)$  commence par les valeurs suivantes : 1, 1, 2, 5, 14, 42. Plus généralement,

$$Cat_n = \frac{1}{n+1} \binom{2n}{n} \quad (1)$$

est le  $n$ -ième nombre de Catalan. Ces nombres énumèrent de multiples modèles combinatoires.

Pour démontrer ce résultat, nous allons utiliser un codage des arbres plan par une fonction, appelée son *contour*. Informellement, on suit la « hauteur » d'une particule qui fait le « contour » de l'arbre de gauche à droite (voir la figure 6 pour un exemple qui sera plus éclairant qu'une définition formelle).

FIGURE 6 – Un arbre plan et son contour



On peut vérifier que ce codage est une bijection entre les arbres plans à  $n + 1$  sommets et des fonctions continues qui augmentent linéairement de  $\pm 1$  (qu'on appellera excursions discrètes) qui sont positives sur  $[0, 2n]$  et nulles en 0 et  $2n$ . Remarquons que chaque arête de l'arbre correspond en fait à deux pas de la marche : un pas ascendant et un pas descendant qui se font « face » sous le graphe de la marche, voir figure 6.

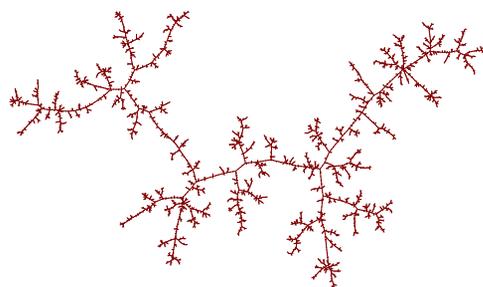
2. Il existe un algorithme encore plus rapide, linéaire en le nombre de sommet, dû à Devroye.

Le fait que le nombre d'excursions discrètes est donné par un nombre de Catalan est classique, et on peut montrer plus généralement (voir par exemple [5, Sec. 6.3]) que

$$\frac{f}{2n+f} \binom{2n+f}{n} \quad (2)$$

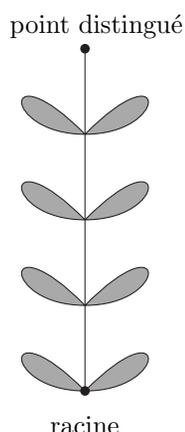
est le nombre de forêts à  $f$  arbres plans ayant  $n$  arêtes.

**Simulation.** Voici une simulation d'un arbre plan aléatoire à 10000 sommets :



Comme pour les arbres de Cayley, nous n'avons pas fait la liste de tous les arbres plans à  $n = 10000$  sommets avant d'en choisir un uniformément au hasard. Grâce à notre codage par la fonction de contour, il nous suffit de simuler une excursion discrète de longueur  $2n$ . Or d'après (1) la probabilité pour qu'une marche de  $2n$  pas à incréments  $\pm 1$  tirée uniformément au hasard parmi les  $2^{2n}$  possibilités soit une excursion est d'ordre  $Cat_n \cdot 2^{-2n} \sim \frac{1}{2\sqrt{\pi n^{3/2}}}$ . En moyenne, il faut donc faire environ  $2\sqrt{\pi n^{3/2}}$  tirages<sup>2</sup> afin d'obtenir une excursion discrète de longueur  $2n$ .

**Application.** La formule explicite (2) permet d'obtenir une information géométrique, à savoir la loi de la hauteur  $\mathcal{H}_n$  d'un sommet choisi uniformément au hasard dans un arbre plan uniforme à  $n$  arêtes. En effet, un arbre à  $n$  arêtes avec un sommet distingué à hauteur  $h \geq 0$  est obtenu à partir d'une forêt de  $2h + 1$  arbres recollés sur une branche de longueur  $h$  (voir figure ci-dessous).



La quantité  $\mathbb{P}(\mathcal{H}_n = h)$  vaut donc

$$\frac{\text{Card}(\{\text{forêt } (2h+1) \text{ arbres à } n-h \text{ arêtes}\})}{(n+1) \cdot \text{Cat}_n} = \frac{2h+1}{2n+1} \frac{\binom{2n+1}{n-h}}{\binom{2n}{n}}$$

L'utilisation de la formule de Stirling montre alors que pour  $x \geq 0$  fixé,  $\mathbb{P}(\mathcal{H}_n > x\sqrt{n}) \rightarrow \exp(-x^2)$  lorsque  $n \rightarrow \infty$ . Nous reconnaissons la loi de Rayleigh (à un facteur de dilatation  $1/\sqrt{2}$  près) apparue dans la Section 2.1. Nous verrons dans la Section 3 que ce n'est pas une coïncidence.

### 2.3 – Bienaymé-Galton-Watson

Les arbres aléatoires plans ou de Cayley sont en réalité des cas particuliers d'un modèle plus général, à savoir celui des arbres de Bienaymé-Galton-Watson (BGW). Il s'agit d'un modèle simple décrivant la généalogie d'une population se reproduisant de manière asexuée qui fut initialement introduit pour étudier la survie des noms de famille nobles anglais. Voir [2] pour une présentation accessible dès les premières années postbac.

Informellement, si  $\mu$  est une probabilité sur  $\mathbb{N}$  (appelée la loi de reproduction), on construit un arbre plan aléatoire issu d'un ancêtre (la racine) et où la probabilité qu'un individu ait  $k$  enfants est  $\mu(k)$ . Dès que la moyenne  $\sum_{k \geq 0} k\mu(k)$  du nombre d'enfants est plus petite ou égale à 1 (et que l'on exclut le cas trivial  $\mu(1) = 1$ ), la lignée s'éteint et l'arbre aléatoire est fini. Quand la moyenne est égale à 1 on dit que la loi de reproduction est *critique*, c'est le cas qui va nous intéresser dans la suite. Il est alors possible de vérifier (voir [2]) que :

- si  $\mu(k) = \frac{1}{2^{k+1}}$  pour  $k \geq 0$ , alors un arbre de BGW de loi de reproduction  $\mu$ , conditionné à

avoir  $n$  sommets, a la même loi qu'un arbre de Catalan uniforme à  $n$  sommets ;

- si  $\mu(k) = \frac{e^{-1}}{k!}$  pour  $k \geq 0$ , alors un arbre de BGW de loi de reproduction  $\mu$ , conditionné à avoir  $n$  sommets, muni d'un étiquetage uniforme des sommets et dont on oublie la structure planaire, a la même loi qu'un arbre de Cayley uniforme à  $n$  sommets.

## 3. Arbres et espaces métriques aléatoires

Dans les deux sections précédentes, nous avons obtenu un théorème limite pour la hauteur d'un point « typique » dans un « arbre aléatoire » (dans deux cadres : Cayley et Catalan). Mais comment faire pour contrôler « globalement » la géométrie, c'est-à-dire les distances entre tous les points en même temps ? Le plus dur est de trouver le bon cadre. C'est au début des années 2000 que Evans, Pitman & Winter, suivis par Duquesne & Le Gall, ont compris que ce cadre existait déjà en géométrie métrique et l'ont importé en probabilité.

### 3.1 – Topologie de Gromov-Hausdorff

L'idée, apparue à la fin des années 1970, due à Edwards et redécouverte par Gromov dans un autre contexte, est de considérer en quelque sorte l'espace de tous les espaces métriques, ou, plus précisément,

$$\mathbb{K} = \left\{ \begin{array}{l} \text{classes d'isométrie d'espaces} \\ \text{métriques compacts non vides} \end{array} \right\},$$

puisque du point de vue géométrique il est impossible de distinguer deux espaces métriques isométriques (en particulier, dans la suite, quand nous parlerons d'un espace métrique, il faudra comprendre implicitement que nous considérons sa classe d'isométrie).

On pourrait penser que cet ensemble est monstrueux et que sa définition même peut poser problème. En réalité, grâce à la condition de compacité imposée sur ses points (i.e. sur les classes d'isométrie d'espaces métriques), cet ensemble est assez « petit », par exemple n'importe quel espace métrique compact peut être vu comme un fermé de  $\ell^\infty(\mathbb{R})$ .

Nous allons maintenant munir  $\mathbb{K}$  d'une distance, connue sous le nom de distance de Gromov-Hausdorff et notée  $d_{GH}$ . Soit  $X$  et  $Y$  deux points de  $\mathbb{K}$ , c'est-à-dire deux (classes d'isométrie d') espaces métriques compacts. L'idée est simple : deux espaces métriques compacts  $X$  et  $Y$  sont identifiés, c'est-à-dire à distance 0 s'il existe une isométrie  $X \rightarrow Y$  et une isométrie  $Y \rightarrow X$ . Plus généralement, la distance de Gromov-Hausdorff entre  $X$  et  $Y$  est plus petite que  $\varepsilon > 0$  si il existe deux  $\varepsilon$ -isométries une de  $X$  vers  $Y$  et une de  $Y$  vers  $X$  où  $\phi : (A, d_A) \rightarrow (B, d_B)$  est une  $\varepsilon$ -isométrie si pour

$$\forall x, x' \in A : |d_B(\phi(x), \phi(x')) - d_A(x, x')| \leq \varepsilon.$$

Cette définition forme bien une distance à laquelle le nom de Hausdorff est attaché à cause de la définition alternative de  $d_{GH}$

$$d_{GH}(X, Y) = 2 \inf\{d_{Haus, E}(X', Y')\}$$

où  $d_{Haus, E}(X', Y')$  est la distance de Hausdorff entre  $X' \subset E$  et  $Y' \subset E$  deux compacts d'un même espace ambiant  $E$  qui sont respectivement isométriques à  $X$  et  $Y$ .

**Théorème 1.** *L'espace  $(\mathbb{K}, d)$  est un espace métrique polonais (i.e. séparable et complet).*

Nous renvoyons le lecteur et la lectrice à [1, Chapitre 7] pour des détails concernant cet espace.

### 3.2 – Espaces métriques aléatoires

Une fois ce cadre posé, un espace (métrique compact) aléatoire n'est rien qu'une variable aléatoire à valeurs dans  $(\mathbb{K}, d_{GH})$ , tout comme un nombre aléatoire est une variable aléatoire à valeurs dans  $\mathbb{R}$ . Par ailleurs, la convergence d'espaces métriques (compacts) aléatoires est la convergence en loi sur  $(\mathbb{K}, d_{GH})$  : rappelons qu'une suite de variables aléatoires  $(X_n)$  à valeurs dans  $\mathbb{K}$  converge en loi vers une variable aléatoire  $X$  si

$$\mathbb{E}[F(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[F(X)],$$

pour toute  $F : \mathbb{K} \rightarrow \mathbb{R}$  continue bornée (par exemple, le cosinus du diamètre de l'espace métrique).

Ce formalisme est bien pratique et permet de donner une définition de l'arbre continu brownien, comme la « limite d'échelle » d'arbres discrets aléatoires renormalisés. En effet, en utilisant le cadre de la Section 2.3, considérons une loi de reproduction  $\mu$  critique et notons  $\mathcal{T}_n$  un arbre de BGW de loi

de reproduction  $\mu$  conditionné à avoir  $n$  arêtes. Cet arbre discret aléatoire, une fois muni de sa distance de graphe peut être vu comme un espace métrique fini (donc compact!), que nous noterons  $(\mathcal{T}_n, d_{gr})$ .

D'après les calculs des deux sections précédentes, nous savons que l'ordre de grandeur des distances dans cet espace métrique aléatoire est d'ordre  $\sqrt{n}$  dans le cas d'une loi de reproduction géométrique ou de Poisson. Il est donc légitime de renormaliser la distance par ce facteur et de considérer

$$(\mathcal{T}_n, \frac{1}{\sqrt{n}} d_{gr}).$$

Cet ordre de grandeur apparaît plus généralement dès que la variance de  $\mu$  est finie (la raison, liée au théorème central limite, sera expliquée un peu plus bas).

**Théorème-Définition 1 (Aldous, Le Gall).** *Soit  $\mu$  une loi de reproduction critique de variance  $\sigma^2$  finie et non nulle. Alors nous avons la convergence en loi suivante au sens de Gromov-Hausdorff*

$$(\mathcal{T}_n, \frac{1}{\sqrt{n}} d_{gr}) \xrightarrow[n \rightarrow \infty]{(d)} (\mathcal{T}, \frac{2}{\sigma} d),$$

où  $(\mathcal{T}, d)$  est un arbre continu compact aléatoire, appelé arbre continu brownien, et dont la loi ne dépend pas de  $\mu$ .

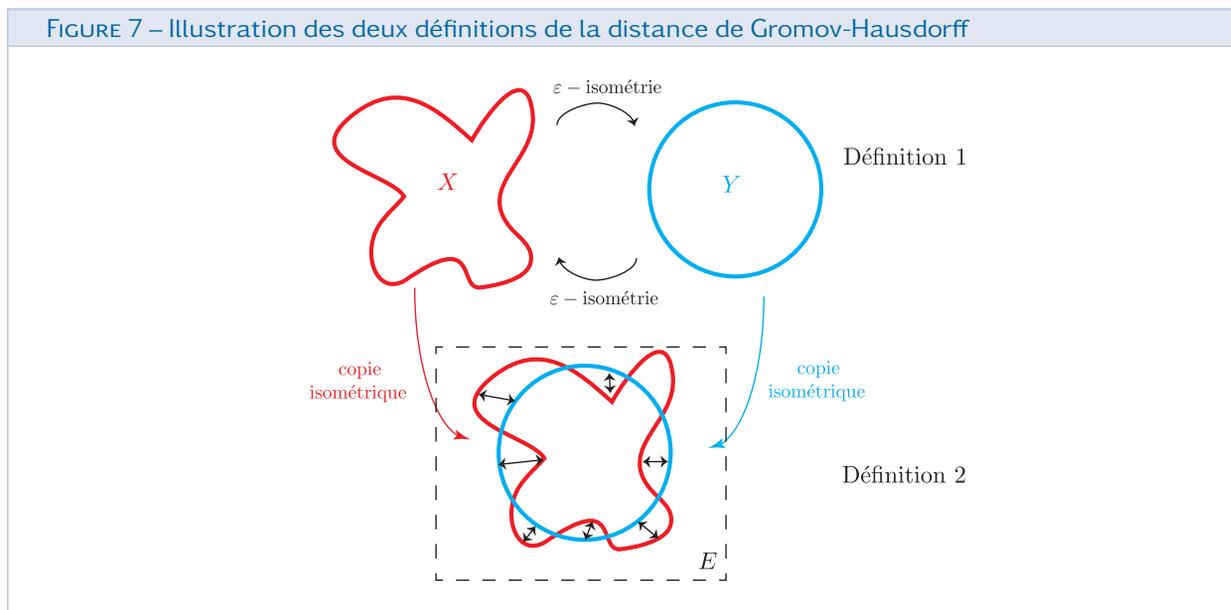
L'arbre continu brownien  $\mathcal{T}$ , appelé fréquemment CRT (pour « continuum random tree ») dans la littérature, est donc un espace métrique aléatoire (par exemple, son diamètre est aléatoire) mais il possède des propriétés « presque sûres » i.e. vraies avec probabilité 1 :

- $\mathcal{T}$  est p.s. un arbre continu, c'est-à-dire un espace métrique compact, géodésique (c'est-à-dire dans lequel deux points quelconques sont reliés par une unique géodésique) et sans cycle.
- pour tout  $x \in \mathcal{T}$ , l'espace  $\mathcal{T} \setminus \{x\}$  a au plus 3 composantes connexes.
- la dimension fractale de  $\mathcal{T}$  est égale à 2.

### 3.3 – Et le brownien dans tout ça ?

**Apparition du mouvement brownien.** À première vue, il n'y a pas grand chose de « brownien » dans la définition de  $\mathcal{T}$ . Pour comprendre d'où vient ce nom, nous allons nous intéresser à la fonction de contour  $C(\mathcal{T}_n) = (C_s(\mathcal{T}_n))_{0 \leq s \leq 2n}$  (voir figure 6 pour un rappel de la fonction de contour). La quantité  $2n$

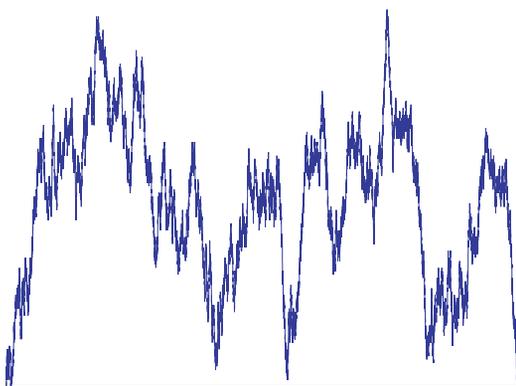
FIGURE 7 – Illustration des deux définitions de la distance de Gromov-Hausdorff



provient du fait que comme chaque arête est visitée deux fois par la fonction de contour, il faut un temps  $2n$  pour faire le contour de l'arbre. Il se trouve que convenablement renormalisée en espace, cette fonction de contour converge. Plus précisément, la convergence

$$\left( \frac{C_s(\mathcal{T}_n)}{\sqrt{n}} : 0 \leq s \leq 1 \right) \xrightarrow[n \rightarrow \infty]{(d)} \left( \frac{2}{\sigma} \mathbf{e}_t : 0 \leq t \leq 1 \right)$$

a lieu en loi dans l'espace  $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ , où  $\mathbf{e}$  est une fonction continue aléatoire, appelée *excursion brownienne*, et qui peut informellement être vue comme un mouvement brownien qui part de 0 au temps 0, reste positif sur l'intervalle de temps  $(0, 1)$  et revient en 0 au temps 1 (voir ci-dessous pour une simulation).



3. Sauf dans le cas des arbres de Catalan, i.e. quand la loi de reproduction est géométrique.

La raison pour laquelle le mouvement brownien apparaît est que, bien qu'en général la fonction de contour ne soit pas une marche aléatoire<sup>3</sup>, elle peut cependant bien être approchée par une marche aléatoire, de sorte que la convergence précédente est une application (dans un cadre conditionné) du théorème de Donsker, selon lequel des marches aléatoires convenablement renormalisées convergent vers l'excursion brownienne (il s'agit de l'extension fonctionnelle du théorème central limite). En particulier, le facteur de renormalisation  $\sqrt{n}$  est le même que dans le théorème central limite grâce à l'hypothèse de variance finie.

**Construction de l'arbre continu brownien.** Il est alors naturel de s'attendre à ce que l'excursion brownienne code, en un certain sens, l'arbre continu brownien. Cette intuition a été formalisée par Duquesne & Le Gall, qui ont mimé la construction d'un arbre à partir de sa fonction de contour. Plus précisément, à toute fonction continue  $f : [0, 1] \rightarrow \mathbb{R}_+$  telle que  $f(0) = f(1) = 0$ , on associe une pseudo-distance sur  $[0, 1]$ , notée  $d_f$ , et définie par

$$d_f(s, t) = f(s) + f(t) - 2 \min_{u \in [s \wedge t, s \vee t]} f(u).$$

Il est aisé de vérifier que  $d_f$  est une pseudo-distance et que les points  $s, t \in [0, 1]$  à distance nulle sont ceux qui se font « face » sous le graphe de  $f$ . On

peut alors considérer la relation d'équivalence sur  $[0, 1]$  obtenue en posant  $s \sim t$  si  $d_f(s, t) = 0$ . Sur l'espace quotient  $[0, 1]/\sim$  la (projection de) la pseudo-distance  $d_f$  est maintenant une distance. Il est possible de vérifier que  $([0, 1]/\sim, d_f)$  est un espace métrique compact, noté  $\mathcal{T}_f$ , qui est un arbre continu. Lorsque la construction précédente est effectuée en partant de l'excursion brownienne, l'arbre aléatoire  $\mathcal{T}_e$  est l'arbre continu brownien  $(\mathcal{T}, d)$  qui apparaît dans le Théorème-Définition 1 (pour être tout à fait précis, Aldous a démontré en 1991 la convergence des fonctions de contour mentionnée plus haut, Le Gall a reformulé la convergence en termes d'espaces métriques compacts aléatoires, voir en particulier [3]).

**Application.** La construction de l'arbre continu brownien à partir de l'excursion brownienne permet (outre d'expliquer son nom) de prouver de nombreuses propriétés de  $\mathcal{T}$ . En particulier, Rényi et Szekeeres ont démontré dans les années 70 que la hauteur maximale  $Ht_n$  d'un arbre de Cayley uniforme à  $n$  arêtes converge en loi  $Ht_n/\sqrt{n} \rightarrow \Theta$  quand  $n \rightarrow \infty$  où la loi  $\Theta$  a pour densité

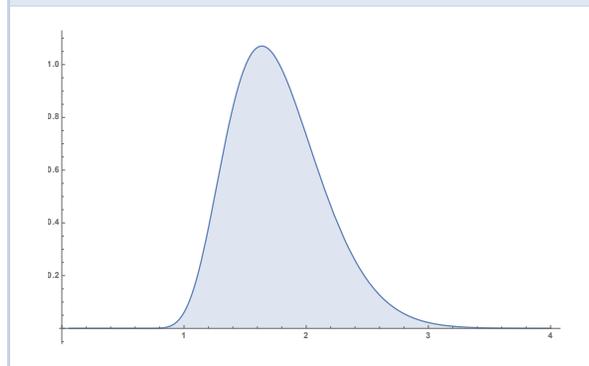
$$4x \sum_{k \geq 1} k^2(2k^2x^2 - 3)e^{-k^2x^2} dx \mathbf{1}_{x>0}.$$

Il était également connu grâce à Chung que le maximum de l'excursion brownienne suivait cette loi. À l'époque, il n'était pas clair de savoir si c'était une coïncidence.

Comme la hauteur d'un arbre est le supremum de sa fonction de contour et que le supremum est une fonctionnelle continue sur  $(\mathcal{C}([0, 1]), \|\cdot\|_\infty)$ , le Théorème-Définition 1 montre que ce n'est pas une coïncidence : intuitivement, l'excursion brownienne toute entière gouverne la géométrie globale de grands aléatoires de Cayley.

La construction précédente permet également de démontrer les propriétés de l'arbre continu brownien évoquées à la fin de la Section 3.2.

FIGURE 8 – Densité de la loi theta, apparaissant comme la loi de la hauteur du CRT ou comme la loi du supremum de l'excursion brownienne



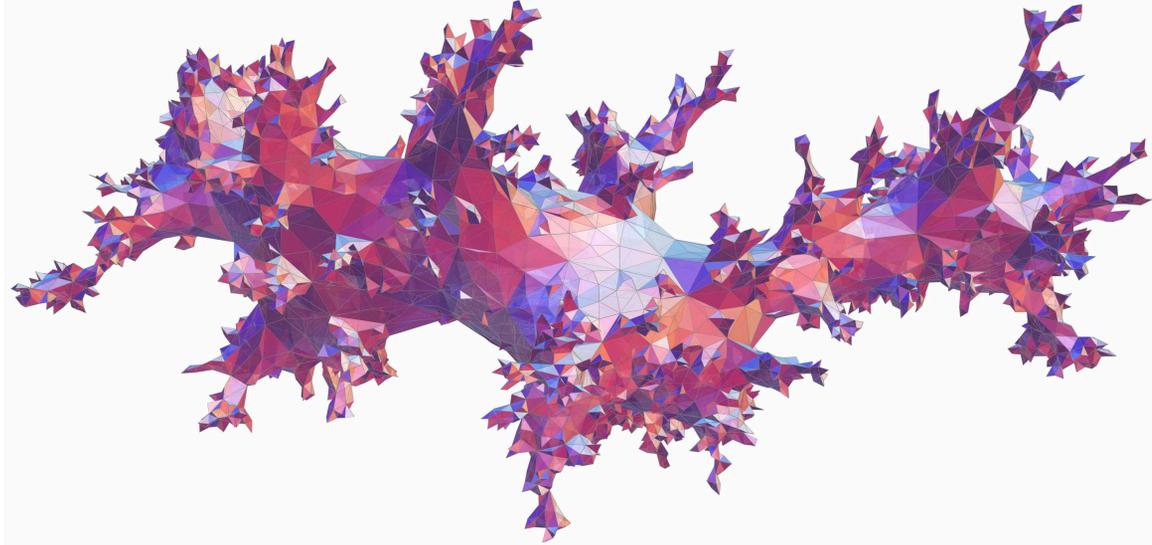
### 3.4 – À quoi cela sert-il ?

Depuis son introduction par Aldous au début des années 90, l'arbre continu brownien est apparu comme une limite universelle de nombreuses structures géométriques arborescentes aléatoires. Nous avons vu le cas des arbres de Bienaymé-Galton-Watson (englobant le cas des arbres de Cayley, et des arbres plans uniformes). Nous pouvons revenir également sur notre exemple introductif des alcanes grâce à un théorème récent de Haas & Miermont : un alcane uniforme à  $n$  arêtes, vu comme un espace métrique converge après remise à l'échelle par  $1/\sqrt{n}$  vers le CRT (à un facteur multiplicatif près).

Le CRT apparaît également comme limite d'échelle de structures qui ne sont pas strictement arborescentes dans le discret mais le deviennent à la limite comme les amas de percolation critique en grande dimension (ou dans le graphe complet). En physique, il est connu sous le nom de « branched polymer ».

Le CRT joue un rôle crucial dans la théorie des cartes aléatoires et de la gravité quantique en dimension 2 car c'est la brique de base pour la construction de la *sphère brownienne*, voir [4] et la figure 9 pour une simulation. Mais le plus important, c'est que 20 ans après sa découverte, le CRT garde encore de nombreux mystères.

FIGURE 9 – Une triangulation aléatoire à 3000 sommets. Sa limite d'échelle, la sphère brownienne est construite à partie de l'arbre continu brownien.



## Références

- [1] D. BURAGO, Y. BURAGO et S. IVANOV. *A course in metric geometry*. 33. Graduate Studies in Mathematics. Providence, RI : American Mathematical Society, 2001, p. xiv+415. ISBN : 0-8218-2129-6.
- [2] I. KORTCHEMSKI. « Arbres et marches aléatoires ». In : *Arbres et marches aléatoires*. Ed. Éc. Polytech., Palaiseau, 2016, p. 1-54.
- [3] J.-F. LE GALL. « Random trees and applications ». *Probability Surveys* (2005).
- [4] J.-F. LE GALL et G. MIERMONT. « Scaling limits of random trees and planar maps ». *Lecture notes for the Clay Mathematical Institute Summer School in Buzios* (July 11 - August 7, 2010).
- [5] J. PITMAN. *Combinatorial stochastic processes*. 1875. Lecture Notes in Mathematics. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard. Berlin : Springer-Verlag, 2006, p. x+256. ISBN : 978-3-540-30990-1; 3-540-30990-X.



### Nicolas CURIEN

Université Paris-Saclay et Institut Universitaire de France  
nicolas.curien@gmail.com

Nicolas Curien est professeur au Laboratoire de mathématique d'Orsay et membre junior de l'IUf. Il travaille sur les propriétés géométriques des grands graphes aléatoires et de leurs limites continues.



### Igor KORTCHEMSKI

CNRS & École polytechnique  
igor.kortchemski@math.cnrs.fr

Igor Kortchemski est chargé de recherche CNRS au CMAP (Centre de Mathématiques Appliquées à l'École polytechnique). Il s'intéresse au comportement asymptotique de grandes structures discrètes aléatoires.