

Introduction au domaine de recherche.
École Normale Supérieure

Limite d'échelle d'arbres de Galton-Watson.

Othmane SAFSAFI
Sous la direction de Nicolas BROUTIN



FIGURE 1 – Un arbre de Galton Watson à 4107 noeuds (image faite avec python).

Résumé

Cette introduction au domaine de recherche est un bref résumé de certains résultats importants dans le domaine des limites d'échelles de graphes aléatoires.

Pour alléger la lecture ce document ne contient pas de démonstrations, cependant tous les théorèmes sont accompagnés de références avec les démonstrations.

Dans la littérature le mot limite d'échelle indique une limite continue et aléatoire d'objets discrets aléatoires remis à l'échelle par une reparamétrisation du temps et une homothétie. Ces objets dépendent généralement de paramètres et on les étudie à un paramètre critique où un changement radical de comportement est observé. En ce sens, le théorème central limite est le premier théorème de limite d'échelle. Rappelons ce théorème :

Théorème 1. *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires iid (indépendantes identiquement distribuées) centrées réduites. Soit $S_n = \sum_{i=1}^n X_i$ on a alors :*

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1),$$

la convergence ici a lieu en loi, ceci veut dire que pour toute fonction continue bornée ϕ , on a :

$$\mathbb{E}\left(\frac{S_n}{\sqrt{n}}\right) \rightarrow \mathbb{E}(\mathcal{N}(0, 1)),$$

$\mathcal{N}(\mu, \sigma^2)$ est la notation standard pour la loi normale de moyenne μ et de variance σ^2 .

Dans ce théorème le fait que les incréments des marches soient centrés représente le paramètre critique, on n'a cependant pas de reparamétrisation du temps car les objets ne sont pas des processus aléatoires. Le théorème de Donsker quant à lui est un théorème de limite d'échelle avec reparamétrisation du temps.

Nous commencerons par une brève présentation des outils nécessaires à la compréhension du sujet, ensuite nous donnerons la limite d'échelle des arbres de Galton-Watson.

La première section présente la notion d'arbre de Galton-Watson ainsi que celle de parcours en profondeur.

La seconde section présente un premier résultat de limite d'échelle relié à une fonction qui code la structure des arbres.

La troisième section introduit le CRT (pour Continuum random tree) qui constituera la limite d'échelle de nos arbres de Galton-Watson.

Enfin la dernière section contient le théorème de convergence vers la limite d'échelle et les idées principales de la démonstration.

Table des matières

1	Préliminaires.	3
1.1	Arbres de Galton-Watson.	3
1.2	Processus reliés aux arbres.	4
2	Limites d'échelles des fonctions codantes.	7
2.1	Les excursions browniennes.	7
2.2	Limites d'échelles des processus reliés aux arbres.	7
3	Les arbres continus.	9
3.1	Arbres en tant qu'espaces métriques compacts.	9
3.2	Le CRT.	9
4	Limite d'échelle des arbres.	10
4.1	La distance de Gromov-Hausdorff.	10
4.2	Limite d'échelle des arbres de Galton Watson en tant qu'espaces métriques.	10
4.3	Conclusion.	10

1 Préliminaires.

Définition d'un graphe.

Soit V un ensemble fini de cardinal $n \in \mathbb{N}$.

Un graphe (non orienté) à n sommets étiqueté par V est un couple (V, E) où E est un ensemble de paires d'éléments de V .

- V (pour vertices) est appelé ensemble de sommets (ou noeuds) et ses éléments sont appelés sommets (ou noeuds) du graphes.
- E (pour edges) est appelé ensemble d'arêtes et ses éléments sont appelés arêtes du graphes.

Définition d'un arbre. Soit (V, E) un graphe étiqueté, un chemin dans le graphe est une suite finie d'arêtes de la forme :

$$(\{a_1, a_2\}, \{a_2, a_3\}, \dots, \{a_{k-1}, a_k\}),$$

où $k \in \mathbb{N}$, si $a_1 = a_k$ alors le chemin est appelé cycle.

La distance de graphe entre deux noeuds a et b est le plus petit nombre d'arêtes dans un chemin qui commence en a et fini en b .

Un arbre étiqueté non ordonné est un graphe (V, E) sans cycles.

Si V est un ensemble ordonné alors on peut considérer que les noeuds de l'arbre sont ordonnés suivant cet ordre, dans ce cas on dit que l'arbre est ordonné.

Soit Φ_V l'ensemble des bijections de V dans lui même, pour $\phi \in \Phi_V$ on peut appliquer ϕ à E pour obtenir un autre ensemble de paires de noeuds, on notera cet ensemble $\phi(E)$.

Notons Ψ_V l'ensemble des ensembles de paires de noeuds de V .

On considère la relation d'équivalence \sim_V sur Ψ_V :

$$(A, B) \in (\Psi_V)^2, A \sim_V B \iff \exists \phi \in \Phi_V, B = \phi(A)$$

Soit \mathbb{E} la classe de E dans le quotient de Ψ_V par la relation d'équivalence \sim_V , (V, \mathbb{E}) est alors un arbre non étiqueté non ordonné.

Si on équipe V d'un ordre δ , alors (V, \mathbb{E}, δ) est un arbre non étiqueté ordonné.

Pour tout les types d'arbres on peut choisir un noeud particulier qu'on appelle racine. Dans ce cas on dit que l'arbre est enraciné.

Comme on vient de le voir il existe plusieurs types différents d'arbres selon qu'ils sont étiquetés ou non, ordonnés (parfois on dit aussi planaire) ou non et enracinés ou non. Il est cependant pratique de s'imaginer un arbre comme étant plongé dans le plan. Pour plus de propriétés des arbres voir [LeGall, 2005] où [Lambert, 2015].

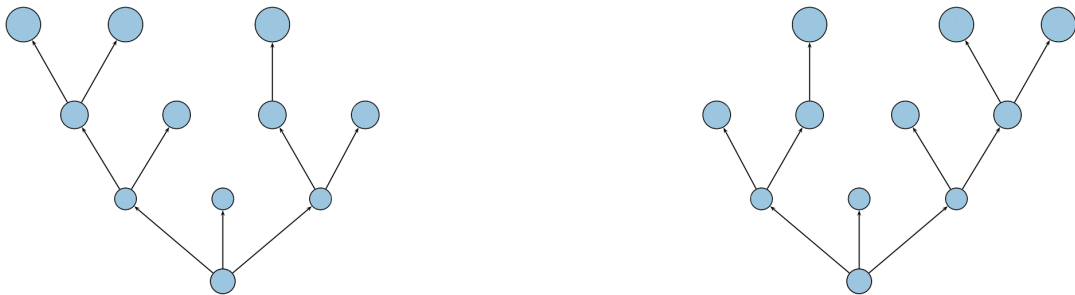


FIGURE 2 – Deux arbres, en tant qu'arbres non ordonnés ce sont les mêmes mais en tant qu'arbres ordonnés ils sont différents. (image faite avec inkscape)

1.1 Arbres de Galton-Watson.

Un arbre de Galton-Watson de loi de reproduction μ est un arbre (ordonné non étiqueté) aléatoire dont chaque noeud a un nombre aléatoire d'enfants suivant la loi μ et indépendamment des autres noeuds.

Définition [Galton and Watson, 1874].

Soit μ une mesure de probabilité portée par \mathbb{N} , soit $(X_i)_{i \in \mathbb{N}}$ une suite iid de variables aléatoires de loi μ . L'arbre aléatoire ordonné non étiqueté tel que si a_i est la classe du i -ème noeud dans l'ordre croissant de l'arbre alors le nombre d'enfants de a_i est X_i est un arbre de Galton-Watson de loi de reproduction μ .

Avec cette définition un arbre de Galton-Watson peut être infini, le théorème suivant va nous donner une condition nécessaire et suffisante pour que l'arbre soit fini.

Théorème 2. *Soit μ une loi de probabilité portée par \mathbb{N} .*

Soit T un arbre de Galton-Watson de loi de reproduction μ , est Z_T son nombre total de noeuds.

Si $\mathbb{E}(\mu) \leq 1$ alors $P(Z_T < \infty) = 1$.

De plus si $\mathbb{E}(\mu) = 1$ alors $E(Z_T) = \infty$.

La démonstration de ce théorème classique peut être trouvée dans n'importe quel cours sur les processus de branchement, par exemple [B.Athreya and Ney, 1972].

Le cas $\mathbb{E}(\mu) = 1$ est appelé cas critique, nous supposons qu'on est dans le cas critique pour le reste de ce document.

L'exploration d'un arbre consiste à donner un ordre de ses noeuds d'une façon qui soit "canonique" pour tout les arbres. Le *DFS* (pour depth first search) ou parcours en profondeur est une manière de faire cette exploration qu'on peut résumer dans l'algorithme suivant :

Soit (V, E) un arbre à n noeuds et ρ sa racine.

Algorithme 1 : Le DFS.

Données : (V, E, ρ)

Résultat : Une liste ordonné constitué des éléments de V .

1 Initialisation : $i = 1, U = [\rho], S = [], T = V \setminus \{\rho\}$;

2 tant que $i \leq n$ faire

3 **si U est vide alors**

4 | déplacer le plus petit élément de T dans U ;

5 | Soit u le dernier élément ajouté à U , déplacer tout les voisins de u qui sont dans T dans U dans l'ordre décroissant.

6 | Déplacer u de U à la queue de S et incrémenter la variable i de 1.

7 Retourner S .

La figure 3 illustre les premières étapes du parcours en profondeur d'un arbre (V, E) . Dans cette figure on a pris V comme étant un sous-ensemble de l'ensemble des mots finis sur \mathbb{N} et l'ordre sur V comme étant l'ordre lexicographique. L'étiquetage est fait de telle sorte que cet ordre coïncide avec l'ordre du *DFS*, cet étiquetage s'appelle étiquetage d'Ulam-Harris-Neveu.

Une fois qu'on a un ordre de parcours des arbres on peut construire plusieurs fonction qui "codent" la structure des arbres. Nous verrons plus tard que la convergence de ces fonctions permet d'obtenir la convergence des arbres en tant qu'espaces métriques.

1.2 Processus reliés aux arbres.

La marche de Lukasiewicz.

La marche de Lukasiewicz est la fonction L (processus dans le cas aléatoire) associée à un arbre (V, E) à $n \in \mathbb{N}$ noeuds allant de $\{0, \dots, n\}$ dans $\{0, \dots, n\}$ telle que :

$$L(k) = L(k - 1) + V(F_{\mathbb{T}}(k - 1)) - 1.$$

Avec $L(0) = 0$, V est la fonction allant de $\{0, \dots, n\}$ dans \mathbb{N} telle que $V(i)$ est le nombre d'enfants du i -ème noeud dans l'ordre du DFS et $F_{\mathbb{T}}$ est la fonction qui donne les noeuds de \mathbb{T} dans l'ordre du *DFS*.

Si on regarde la figure 4, la marche de Lukasiewicz donne le nombre de noeuds rouges à chaque étape.

On a toujours $L(n) = -1$.

La figure 4 est un exemple de marche de Lukasiewicz.

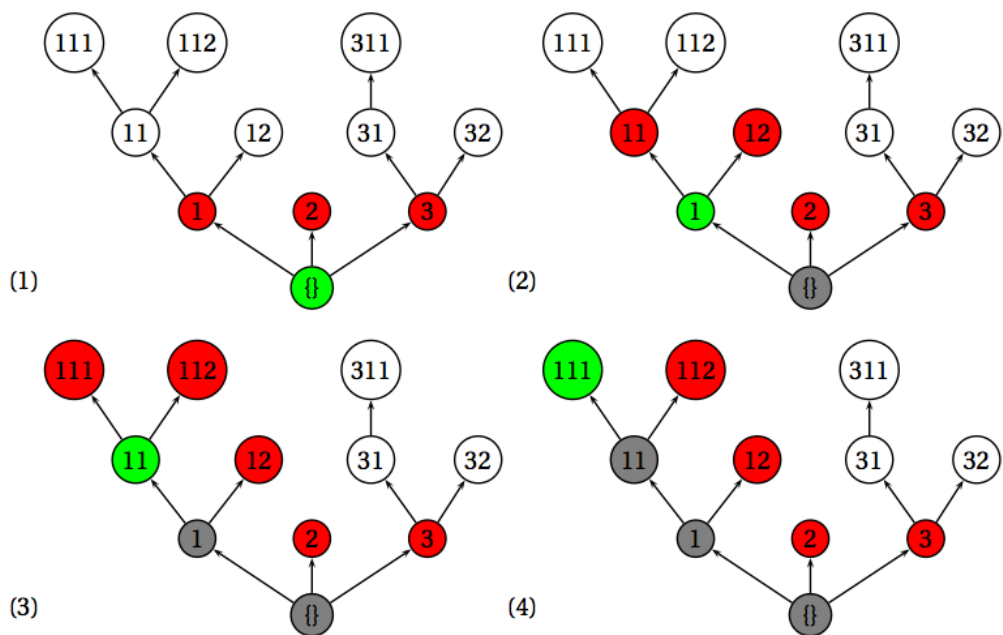


FIGURE 3 – Les 4 premières étapes du DFS d'un arbre, le noeud vert est celui qu'on explore, les noeuds en rouge sont les noeuds dans l'ensemble U , ceux en blanc sont dans T et ceux en gris sont dans S . (image faite sous TikZ)

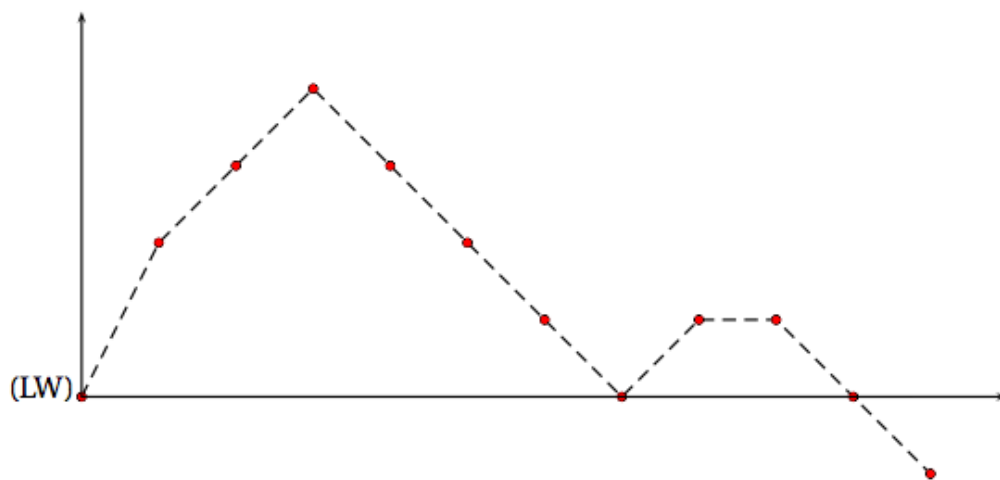


FIGURE 4 – La marche de Lukasiewicz associée à l'arbre de la figure 3

Le DFSR (pour depth first search with redundancy) est une version un peu modifiée du DFS, on le définit directement comme une fonction. Cette appellation est non-standard, le lecteur trouvera selon les sources le DFS défini comme étant le DFSR ou comme nous l'avons défini dans ce document.

Définition

Soit (V, E) un arbre ordonné à n noeuds et enraciné en ρ , le DFSR de cet arbre est la fonction dr allant de $\{0, \dots, 2n - 2\}$ dans V telle que :

$$dr(0) = \rho$$

Étant donné $dr(\{0, 1, \dots, i\})$, $dr(i + 1)$ est l'enfant le plus petit dans l'ordre de l'arbre de $dr(i)$ qui n'appartient pas à $dr(\{0, 1, \dots, i\})$. Si $dr(i)$ n'a pas de tels enfants alors $dr(i + 1)$ est le parent de $dr(i)$.

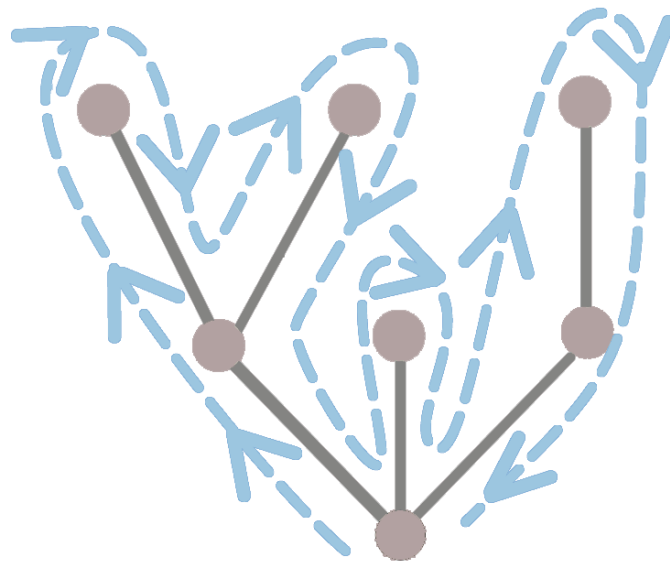


FIGURE 5 – La DFSR de l'arbre consiste à noter la suite des noeuds rencontrés dans le parcours en bleu ciel. (image faite avec Inkscape)

Le processus de contour.

Soit un arbre (V, E) enraciné à n noeuds. La fonction (ou processus dans le cas aléatoire) de contour associée à un arbre (V, E) à n noeuds est la fonction f de $\{0, \dots, 2n - 2\}$ dans \mathbb{N} telle que :

$$f(i) = h(dr(i)),$$

où dr est le DFSR de l'arbre et h est la fonction qui à chaque noeud donne sa distance à la racine (aussi appelée fonction hauteur).

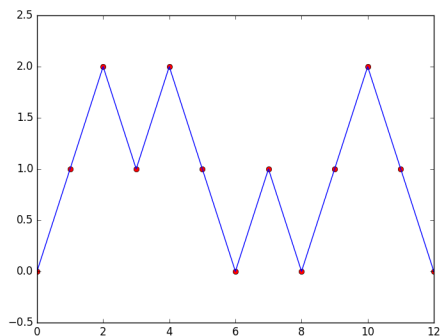
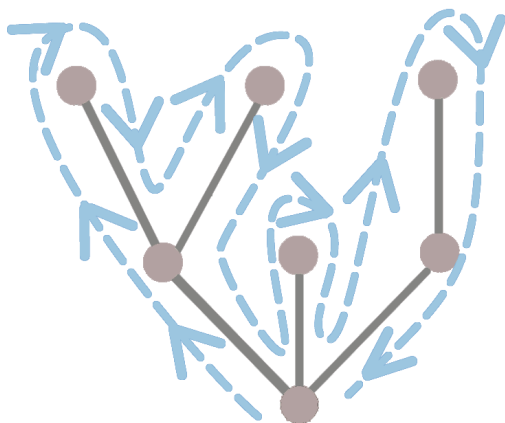


FIGURE 6 – Le processus de contour associé à l’arbre de gauche. (image faite avec python)

2 Limites d’échelles des fonctions codantes.

2.1 Les excursions browniennes.

Le mouvement brownien standard (dans \mathbb{R}) est le seul (à indistinguabilité près) processus stochastique à temps continu W ayant les propriétés suivantes :

- $W(0) = 0$
- $W(t)$ est à accroissements indépendants.
- $W(t)$ est continu presque sûrement.
- $W(t) - W(s) \sim \mathcal{N}(0, t - s)$, $0 \leq s \leq t$.

Définition.

Soit $(W(t))_{t \in [0,1]}$, le processus définit par :

$$b(t) = W(t) - tW(1),$$

b est appelé pont brownien.

Remarquons qu’on a $b_0 = b_1 = 0$ et que b est un processus gaussien.

De manière informelle un pont brownien est un mouvement brownien conditionné à avoir $W_1 = 0$.

Définition

Soit $(b(t))_{t \in [0,1]}$ un pont brownien, le processus $(e(t))_{t \in [0,1]}$ définit par :

$$e(t) = b(\tau_m + t \bmod 1) - b(\tau_m),$$

où τ_m est le point (unique presque sûrement) où le minimum de $(b(t))_{t \in [0,1]}$ est atteint est appelé excursion brownienne.

De manière informelle une excursion brownienne $(e_t)_{t \in [0,1]}$ standard est un mouvement brownien sur $[0, 1]$ conditionné à rester positif et à avoir $W_1 = 0$. La figure 7 montre la construction de l’excursion brownienne à partir d’un pont brownien.

Pour plus d’informations sur cette définition voir Vervaat [1979].

2.2 Limites d’échelles des processus reliés aux arbres.

Théorème 3 (Donsker [1952]). Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires iid d’espérance 0 et de variance 1. On définit :

$$W^n(t) = \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{[nt]} X_k + (nt - [nt])X_{[nt]+1} \right).$$

On a alors la convergence suivante :

$$W^{(n)} \xrightarrow{\text{loi}} W,$$

où W est un mouvement brownien standard sur $[0, 1]$, la convergence a lieu en loi et dans l'espace des fonctions continues sur $[0, 1]$ muni de la distance issue de la norme uniforme.

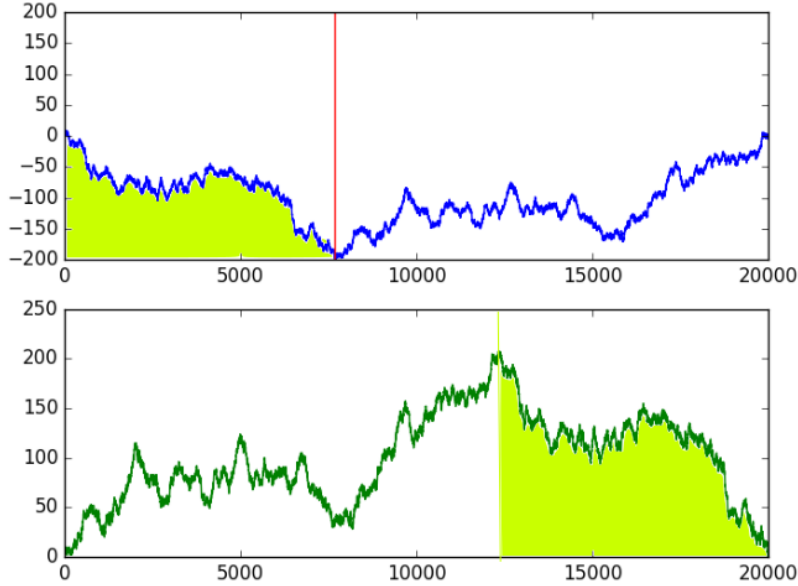


FIGURE 7 – Un pont brownien aléatoire transformé en excursion brownienne . La ligne rouge indique le minimum où la "coupe" est faite, nous avons colorié le bloc déplacé horizontalement. (image faite avec python et inkscape)

Théorème 4 (Donsker [1952], E.Csaki and S.G.Mohanty [1981]). Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires iid d'espérance 0 et de variance 1, supposons de plus que $\mathbb{P}(S_n = 0) > 0$ pour tout n . On définit :

$$W^n(t) = \frac{1}{\sqrt{n}} \left(\sum_{k=1}^{[nt]} X_k + (nt - [nt])X_{[nt]+1} \right).$$

Soit $(e_t^{(n)}, t \in [0, 1])$ $W^{(n)}$ conditionnée à rester positive et à avoir $W_1^{(n)} = 0$, on a alors la convergence suivante :

$$e^{(n)} \xrightarrow{\text{loi}} e,$$

où e est l'excursion brownienne standard et la convergence a lieu en loi et dans l'espace des fonctions continues sur $[0, 1]$ muni de la distance issue de la norme uniforme.

En remarquant que la marche de Lukasiewicz vérifie les conditions de ce théorème à part qu'elle termine en -1 et que ceci n'a aucune incidence sur la limite car on divise par \sqrt{n} , on obtient directement le corollaire suivant.

Corollaire 4.1 (Aldous [1993]). Soit $(\mathbb{T}_n)_{n \in \mathbb{N}}$ une suite d'arbres de Galton-Watson de loi de reproduction μ , $\mathbb{E}(\mu) = 1$ et $\sigma^2 = \text{var}(\mu) < \infty$. Supposons aussi que pour tout n , \mathbb{T}_n est conditionnée à avoir $n \in \mathbb{N}$ noeuds et que cet événement est de probabilité positive.

Soit L_n la marche de Lukasiewicz correspondant à \mathbb{T}_n , et \mathbb{L}_n l'interpolation continue de cette marche sur $[0, 1]$. On a alors :

$$\left(\frac{\mathbb{L}_n(t)}{\sqrt{n\sigma}} \right)_{t \in [0,1]} \xrightarrow{\text{loi}} (e(t))_{t \in [0,1]},$$

où e est l'excursion brownienne standard et la convergence a lieu en loi et dans l'espace des fonctions continues sur $[0, 1]$ muni de la distance issue de la norme uniforme.

3 Les arbres continus.

3.1 Arbres en tant qu'espaces métriques compacts.

Pour pouvoir démontrer la convergence des arbres de Galton-Watson vers des arbres "continus" nous devons d'abord donner une définition d'arbre dans ce cadre.

Définition.

Un \mathbb{R} arbre est un espace métrique complet \mathbb{T} vérifiant les propriétés suivantes :

1. *Unicité des géodésiques* : Pour tout $x, y \in \mathbb{T}$, il existe une unique isométrie $\phi_{x,y} : [0, d(x, y)] \rightarrow \mathbb{T}$ telle que $\phi_{x,y}(0) = x$ et $\phi_{x,y}(d(x, y)) = y$.
La géodésique $\phi_{x,y}([0, d(x, y)])$, aussi appelée arc, est notée $[[x, y]]$.
2. *Pas de cycles* : Pour toute injection continue $\psi : [0, 1] \rightarrow \mathbb{T}$ on a $\psi([0, 1]) = [[\psi(0), \psi(1)]]$.

Si on choisit un point particulier de l'espace métrique comme racine on dira que le \mathbb{R} arbre est enraciné.

Remarquons qu'un arbre discret peut être vu comme un \mathbb{R} arbre compact, il suffit pour cela de considérer les arêtes comme un ensemble continu de points et de prendre comme distance la distance de graphe. Pour un \mathbb{R} arbre T et a un entier positif, on notera aT l'arbre dans lequel les distances entre les points ont été multipliées par a .

3.2 Le CRT.

Proposition (Aldous [1993]). Soit h une fonction continue à support compacte $[0, \sigma]$. Soit d_h la pseudo-distance définie comme suit :

$$\forall s, t \in [0, \sigma] \quad d_h(s, t) = h(s) + h(t) - 2 \inf_{[s \vee t, s \wedge t]} h.$$

Soit \sim_h la relation d'équivalence induite par d_h sur $[0, \sigma]$. Alors le quotient $\mathbb{T}_h := [0, \sigma] / \sim_h$ forme un \mathbb{R} arbre compact avec la distance induite par d_h .

Le *CRT* (pour continuum random tree) est le \mathbb{R} arbre T_{2e} codé par $2e$ où e est l'excursion brownienne standard.

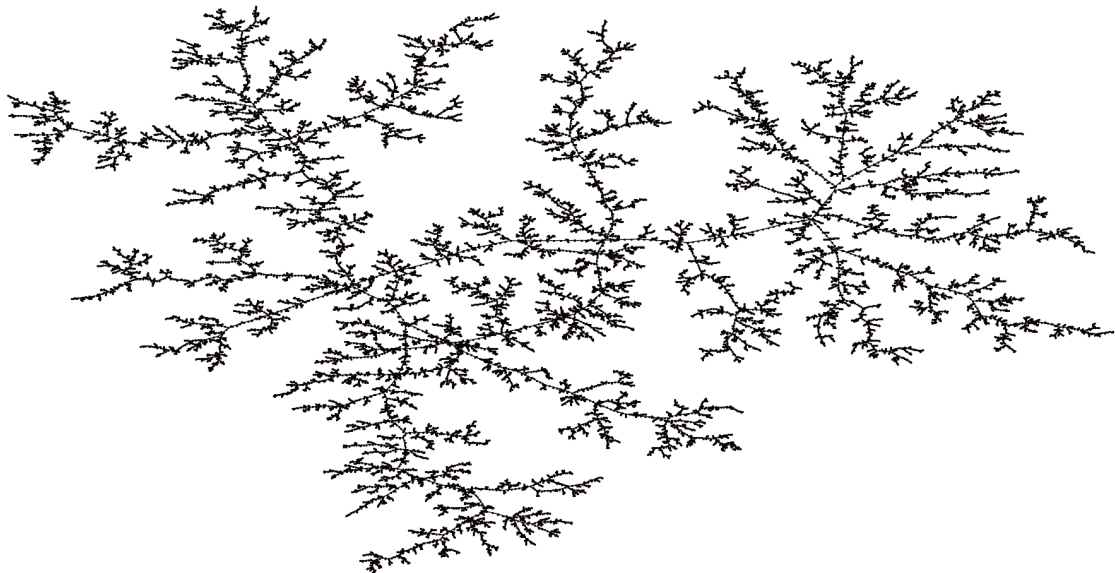


FIGURE 8 – Un arbre aléatoire uniforme ayant 22274 noeuds, approximant le CRT. (image faite avec python)

4 Limite d'échelle des arbres.

4.1 La distance de Gromov-Hausdorff.

Maintenant que nos arbres sont vus comme des espaces métriques il nous faut définir une distance entre ces différents arbres.

Définition.

Soit (M, δ) un espace métrique, la distance de Hausdorff d_H entre deux sous-espaces compacts K, K' de M est :

$$d_H(K, K') = \inf\{\epsilon > 0 : K \subseteq F_\epsilon(K'), K' \subseteq F_\epsilon(K)\},$$

où $F_\epsilon(K) := \{x \in M : \delta(x, K) \leq \epsilon\}$. Supposons maintenant que X et X' sont deux \mathbb{R} arbres compacts (pas forcément dans le même grand espace) et enracinés respectivement en ρ et ρ' .

La distance de Gromov-Hausdorff entre X et X' est :

$$d_{GH}(X, X') = \inf\{\max(d_H(\phi(X), \phi'(X')), \delta(\rho, \rho'))\},$$

où l'infimum est pris sur tous les choix d'espaces métriques (M, δ) et tous les plongements isométriques (qui préservent toutes les distances) $\phi : X \rightarrow M$ et $\phi' : X' \rightarrow M$.

L'ensemble des classes d'isométrie de \mathbb{R} arbres avec la distance de Gromov-Hausdorff est complet et séparable (polonais).

Le théorème suivant permet de passer de la convergence des fonctions de contour à celle des arbres munis de la distance de Gromov-Hausdorff.

Théorème 5 (LeGall [2005]). *Soit $h_1, h_2 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ deux fonctions continues à support compact $h_1(0) = h_2(0) = 0$, alors :*

$$d_{GH}(\mathbb{T}_{h_1}, \mathbb{T}_{h_2}) \leq 2d_\infty(h_1, h_2).$$

d_∞ est la distance correspondant à la norme infinie.

4.2 Limite d'échelle des arbres de Galton Watson en tant qu'espaces métriques.

En comparant la marche de Lukasiewicz à la fonction de contour on obtient le résultat suivant.

Théorème 6 (Aldous [1997]). *Si c_n est l'interpolation continue du processus de contour lié à l'arbre \mathbb{T}_n (avec les notations du corollaire 4.1) on a :*

$$\frac{c_n(nt)}{\sqrt{n}} \underset{t \in [0,2]}{\text{loi}} \xrightarrow{\quad} \frac{2}{\sigma} e(t/2),$$

la convergence ici est la convergence en loi dans l'espace des fonctions continues.

Avec ceci on peut enfin énoncer le théorème principal sur la limite d'échelle des arbres de Galton-Watson.

Théorème 7 (Aldous [1997], LeGall [2010]). *Soit $(\mathbb{T}_n)_{(n \in \mathbb{N})}$ une suite d'arbres de Galton-Watson avec comme loi de reproduction μ , $\mathbb{E}(\mu) = 1$ et $\sigma^2 = \text{var}(\mu) < \infty$. Pour tout n , \mathbb{T}_n est conditionné à avoir $n \in \mathbb{N}$ noeuds et on suppose que ce conditionnement a un sens.*

Alors $\frac{\sigma}{\sqrt{n}} \mathbb{T}_n$ converge en loi vers le CRT suivant la distance de Gromov-Hausdorff.

4.3 Conclusion.

Le domaine des limites d'échelles est riche et utilise plusieurs types de techniques différentes comme la combinatoire discrète, la théorie des martingales ou encore la topologie.

Il y a d'autres limites d'échelles connues. Cependant, même si ce domaine a suscité et continue de susciter beaucoup d'attention il reste plusieurs questions ouvertes. On ne connaît par exemple pas les limites d'échelles de la plupart des modèles de graphes aléatoires existants.

Bibliographie

- Jean-François LeGall. Random trees and applications. *Probability Surveys*, 2 :245–311, 2005.
- Amaury Lambert. *Probabilistic Models for the (sub)Tree(s) of Life*. XIX Escola Brasileira de Probabilidade, 2015.
- Francis Galton and Reverant Watson. On the probability of the extinction of families. <http://galton.org/essays/1870-1879/galton-1874-jaiqi-family-extinction.pdf>, pages 138–144, 1874.
- Krishna B.Athreya and Peter E. Ney. *Branching Processes*. Springer, 1972.
- Wim Vervaat. A relation between brownian bridge and brownian excursion. *The annals of probability*, 7 :143–149, 1979.
- Monroe D. Donsker. Justification and extension of doob’s heuristic approach to the kolmogorov-smirnov theorems. *The Annals of Mathematical Statistics*, 23 :277–281, 1952.
- E.Csaki and S.G.Mohanty. Excursion and meander in random walk. *The canadian journal of statistics*, 9(1) :57–70, 1981.
- David Aldous. The continuum random tree iii. *The annals of probability*, 21 :248–289, 1993.
- David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *The annals of probability*, 25 :812–854, 1997.
- Jean-François LeGall. Itô’s excursion theory and random trees. *Stochastic Processes and their Applications*, 120 :721–749, 2010.