Arbres aléatoires

GAGNAIRE Jules, TRONCH Alix

26 avril 2025

1. Notion d'arbre
2. Arbre de Galton Watson
3. Arbres réels

1. Notion d'arbre
   - Définition
   - Fonction de contour et de hauteur
   - Chemin de Lukasiewicz
2. Arbre de Galton Watson
3. Arbres réels

1. Notion d'arbre
   - Définition
   - Fonction de contour et de hauteur
   - Chemin de Lukasiewicz

2. Arbre de Galton Watson

3. Arbres réels

Définition d'un arbre

Définition

Un **arbre enraciné t** est un sous-ensemble fini de $\mathscr{U} := \bigcup_{k=0}^{\infty} \mathbb{N}^k$ tel que :

1. $\varnothing \in \mathbf{t}$.
2. $u \in \mathbf{t} \backslash \{\varnothing\} \Rightarrow \pi(u) \in \mathbf{t}$. Où $\pi(u)$ est le parent de u.
3. Pour chaque $u \in \mathbf{t}$, il existe un entier $k_u(\mathbf{t}) \geq 0$ tel que, pour chaque $j \in \mathbb{N}$, $uj \in \mathbf{t}$ si et seulement si $1 \leq j \leq k_u(\mathbf{t})$.

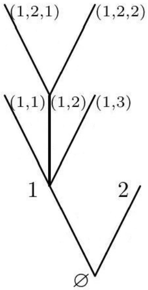

Figure 1 – Arbre $(\varnothing, 1, (1,1), (1,2), (1,2,1), (1,2,2), (1,3), 2)$.

1. Notion d'arbre
   - Définition
   - Fonction de contour et de hauteur
   - Chemin de Lukasiewicz
2. Arbre de Galton Watson
3. Arbres réels

Fonction de contour

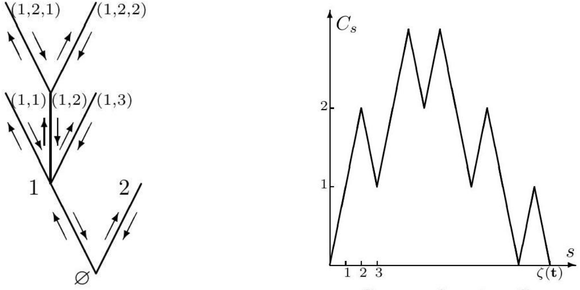

Figure 2 – Illustration de la **fonction de contour** $\zeta(\mathbf{t}) := 2(\#(\mathbf{t})-1)$.

Fonction de hauteur

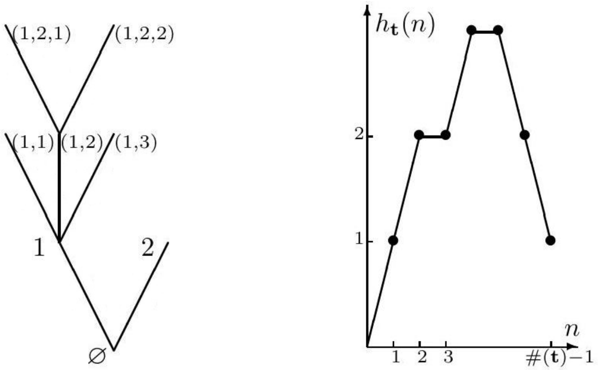

Figure 3 – Illustration de la **fonction de hauteur**.

1 Notion d’arbre
Définition
Fonction de contour et de hauteur
Chemin de Lukasiewicz

2 Arbre de Galton Watson

3 Arbres réels

Les **fonctions de contour et de hauteur caractérisent un arbre**, nous allons voir une autre caractérisation possible :

Notation

On note $\mathscr{S}$ l'ensemble de toutes les suites finies d'entiers positifs $m_1, \ldots, m_p$ telles que :

- $m_1 + m_2 + \cdots + m_i \geq i$, pour $i < p$
- $m_1 + m_2 + \cdots + m_p = p - 1$.

Proposition

L'application

$$\Phi : \mathbf{t} \longrightarrow \left(k_{u_0}(\mathbf{t}), k_{u_1}(\mathbf{t}), \ldots, k_{u_{\#(\mathbf{t})-1}}(\mathbf{t})\right)$$

définit une bijection de $\mathbf{A} = \{\text{arbres ordonnés enracinés}\}$ sur $\mathscr{S}$.

Un exemple

Regardons cette propriété sur un exemple, reprenons l'arbre $(\varnothing, 1, (1,1), (1,2), (1,2,1), (1,2,2), (1,3), 2)$:

$$\Phi(\mathbf{t}) = (2, 3, 0, 2, 0, 0, 0, 0)$$

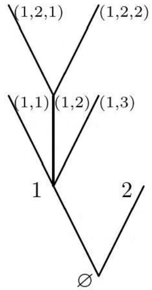

Figure 4 – Reconstruction de l'arbre.

Chemin de Lukasiewicz

Définition

Soit $\mathbf{t} \in \mathbf{A}$ et $p = \#(\mathbf{t})$. Soit $\left(m_1, \ldots, m_p\right) = \Phi(\mathbf{t})$, considérons

$$x_n = \sum_{i=1}^{n} (m_i - 1), \quad 0 \leq n \leq p$$

qui satisfait les propriétés suivantes :

- $x_0 = 0$ et $x_p = -1$.
- $x_n \geq 0$ pour tout $0 \leq n \leq p-1$.
- $x_i - x_{i-1} \geq -1$ pour tout $1 \leq i \leq p$.

Une telle suite est appelée un **chemin de Lukasiewicz**.

Chemin de Lukasiewicz et fonction de hauteur

Proposition

La fonction hauteur $h_{\mathbf{t}}$ d'un arbre $\mathbf{t}$ est liée au chemin de Lukasiewicz de $\mathbf{t}$ par

$$h_{\mathbf{t}}(n) = \#\left\{ j \in \{0, 1, \ldots, n-1\} : x_j = \inf_{j \leq \ell \leq n} x_\ell \right\}$$

pour tout $n \in \{0, 1, \ldots, \#(\mathbf{t}) - 1\}$.

Un exemple

Reprenons le même arbre, $\Phi(\mathbf{t}) = (2,3,0,2,0,0,0,0)$, le chemin de Lukasiewicz est alors $(0,1,3,2,3,2,1,0,-1)$.

Et par exemple :

$$h_{\mathbf{t}}(4) = \#\left\{ j \in \{0,1,2,3\} : x_j = \inf_{j \leq \ell \leq 4} x_\ell \right\} = 2$$

1 Notion d'arbre

2 Arbre de Galton Watson
Définitions et propriétés
Convergence vers un mouvement Brownien
Arbre de Galton Watson à progéniture fixé

3 Arbres réels

1 Notion d'arbre

2 Arbre de Galton Watson
Définitions et propriétés
Convergence vers un mouvement Brownien
Arbre de Galton Watson à progéniture fixé

3 Arbres réels

Rappel de la définition d'un processus de Galton Watson

Définition

Un **processus de Galton-Watson** est un processus stochastique discret $\{Z_n\}_{n\geq 0}$ défini par :

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}, \quad n \geq 0,$$

où :

- Z_n représente le nombre d'individus à la n-ième génération.
- Z_0 est égal au nombre d'individus de la génération initiale.
- Les $X_{n,i}$ sont des v.a. i.i.d., représentant le nombre de descendants produits par l'individu i de la génération n.
- La distribution des $X_{n,i}$, appelée **loi de reproduction**, est notée μ.

Le processus se termine lorsque $Z_n = 0$.

Définition d'un arbre de Galton Watson

- μ mesure de probabilité sur $\mathbb{N}$ critique ou sous critique (i.e. $\sum_{k=0}^{\infty} k\mu(k) \leq 1$).
- $\{K_u | u \in \mathscr{U}\}$ une collection de v.a. i.i.d. .

$$\theta := \left\{ u = (u^1, \ldots, u^n) \in \mathscr{U} \mid u^j < K_{(u^1,\ldots,u^{j-1})}, \forall j \in \{1, \ldots, n\} \right\}$$

Proposition

θ est presque sûrement un arbre. De plus, si

$$Z_n = \#\{u \in \theta : |u| = n\}$$

alors $\{Z_n, n \geq 0\}$ est un processus de Galton-Watson avec loi de reproduction μ et une valeur initiale $Z_0 = 1$. Un tel arbre est appelé **arbre de Galton Watson**.

Proposition

Soit θ un arbre de Galton-Watson associé à la loi μ. Alors

$$\Phi(\theta) \stackrel{(d)}{=} (M_1, M_2, \ldots, M_T)$$

où les variables aléatoires $M_1, M_2, \ldots$ sont indépendantes et suivent la loi μ, et

$$T = \inf\{n \geq 1 : M_1 + \cdots + M_n < n\}.$$

Remarque

- $k_u(\theta) = K_u$.
- Si on écrit $\theta = \{U_0, ..., U_{\#\theta-1}\}$ on a $\phi(\theta) = \left(K_{U_0}, ..., K_{U_{\#\theta-1}}\right)$

Corollaire

Soit $\{S_n, n \geq 0\}$ une marche aléatoire sur $\mathbb{Z}$ avec une valeur initiale S_0 et une distribution de sauts $\nu(k) = \mu(k+1)$ pour $k \geq -1$. On pose

$$T = \inf\{n \geq 1 : S_n = -1\}$$

Alors, le chemin de Lukasiewicz d'un μ-arbre de Galton-Watson θ a la même distribution que $(S_0, S_1, \ldots, S_T)$. En particulier, $\#(\theta)$ et T ont la même distribution.

1 Notion d'arbre

2 Arbre de Galton Watson
Définitions et propriétés
Convergence vers un mouvement Brownien
Arbre de Galton Watson à progéniture fixé

3 Arbres réels

Processus de hauteur

- Fixons une loi de reproduction critique μ avec une variance finie $\sigma^2 > 0$.
- $\theta_1, \theta_2, \ldots$ une suite de μ-arbres de Galton-Watson indépendants. À chaque θ_i, nous associons sa fonction de hauteur $\left(h_{\theta_i}(n), 0 \leq n \leq \#(\theta_i) - 1\right)$.

Définition

On définie le **processus de hauteur** $(H_n, n \geq 0)$ en concaténant les fonctions $h_{\theta_1}, h_{\theta_2}, \ldots$

Proposition

Pour tout $n \geq 0$, nous avons

$$H_n = \#\left\{k \in \{0, 1, \ldots, n-1\} : S_k = \inf_{k \leq j \leq n} S_j\right\} \tag{1}$$

où $(S_n, n \geq 0)$ est une marche aléatoire avec la distribution décrite dans le corollaire.

Idées de la démonstration

- Utiliser le corollaire précédent.
- Utiliser la proposition reliant la hauteur et le chemin de Lukasiewicz.

Définition

Un **mouvement Brownien réfléchi** (partant de l'origine) est la valeur absolue d'un mouvement brownien standard partant de l'origine.

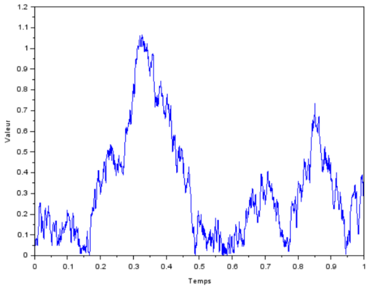

Figure 5 – Mouvement Brownien réfléchi.

Le gros théorème!

Théorème

Nous avons

$$\left(\frac{1}{\sqrt{p}}H_{[pt]}, t \geq 0\right) \xrightarrow[p\to\infty]{(\text{ d})} \left(\frac{2}{\sigma}\gamma_t, t \geq 0\right)$$

où γ est un mouvement Brownien réfléchi.

Corollaire

Nous avons

$$\left(\frac{1}{\sqrt{p}}C_{2pt}, t \geq 0\right) \xrightarrow[p\to\infty]{(\text{d})} \left(\frac{2}{\sigma}\left|\beta_t\right|, t \geq 0\right).$$

Idées de la démonstration

Notons

$$M_n = \sup_{0 \le k \le n} S_k, \quad I_n = \inf_{0 \le k \le n} S_k.$$

Deux étapes principales :

1. La convergence des finis-dimensionels
2. La convergence fonctionelle

Pour le premier point on aura besoin de :

Lemme

Nous avons

$$\frac{H_n}{S_n - I_n} \xrightarrow[n\to\infty]{\mathbb{P}} \frac{2}{\sigma^2},$$

Convergence des finis-dimensionels

- Théorème de donsker donne

$$\left(\frac{1}{\sqrt{p}}S_{[pt]}, t \geq 0\right) \xrightarrow[p\to\infty]{(\text{ d})} (\sigma B_t, t \geq 0)\,. \tag{2}$$

- D'après (2), nous avons, pour tout choix de $0 \leq t_1 \leq t_2 \leq \cdots \leq t_m$,

$$\frac{1}{\sqrt{p}}\left(S_{[pt_1]} - I_{[pt_1]}, \ldots, S_{[pt_m]} - I_{[pt_m]}\right) \xrightarrow[p\to\infty]{(\text{d})} \sigma\left(B_{t_1} - \inf_{0\leq s\leq t_1} B_s, \ldots, B_{t_m} - \inf_{0\leq s\leq t_m} B_s\right).$$

- Par le lemme on a :

$$\frac{1}{\sqrt{p}}\left(H_{[pt_1]}, \ldots, H_{[pt_m]}\right) \xrightarrow[p\to\infty]{(\text{d})} \frac{2}{\sigma}\left(B_{t_1} - \inf_{0\leq s\leq t_1} B_s, \ldots, B_{t_m} - \inf_{0\leq s\leq t_m} B_s\right)$$

- Un théorème de Lévy énonce que le processus

$$\gamma_t = B_t - \inf_{0\leq s\leq t} B_s$$

est un mouvement brownien réfléchi.

Convergence fonctionelle

On va supposer que μ a un petit moment exponentiel (i.e. $\sum_{k=0}^{\infty} e^{\lambda k}\mu(k) < \infty$).

- (M_n, K_n) a la même distribution que $(S_n - I_n, H_n)$
- Soit $A \geq 1$ un entier fixé. On peut montrer que pour p suffisamment grand,

$$P\left[\sup_{0\leq t\leq A}\left|S_{[pt]} - I_{[pt]} - \frac{\sigma^2}{2}H_{[pt]}\right| > (Ap)^{\frac{3}{8}}\right] < Ap\exp\left(-(Ap)^{\varepsilon'}\right).$$

TENSION!!

- Ce qui permet d'obtenir le théorème.

1 Notion d'arbre

2 Arbre de Galton Watson
Définitions et propriétés
Convergence vers un mouvement Brownien
Arbre de Galton Watson à progéniture fixé

3 Arbres réels

Notations et conventions

- Nous supposons que μ a un petit moment exponentiel.
- Pour chaque $p \geq 1$, nous désignons par $\theta^{(p)}$ un arbre de Galton-Watson μ conditionné pour avoir exactement p éléments.
- Nous désignons par $\left(H_k^{(p)}\right)_{0 \leq k \leq p}$ le processus de hauteur de $\theta^{(p)}$.

Convergence des fonctions de hauteurs conditionnées

Définition

Une **excursion Brownienne normalisée** $(\mathbf{e}_t)_{0\leq t\leq 1}$ est une excursion Brownienne conditionnée pour avoir une longueur égale à 1.

Théorème

Nous avons :

$$\left(\frac{1}{\sqrt{p}}H^{(p)}_{[pt]}, 0\leq t\leq 1\right) \xrightarrow[p\to\infty]{(\mathrm{d})} \left(\frac{2}{\sigma}\mathbf{e}_t, 0\leq t\leq 1\right).$$

Idées de la démonstration

- $T_1 := \inf\{n \geq 1 : H_n = 0\} = \inf\{n \geq 0 : S_n = -1\}$.
- On peut montrer que $P(T_1 = p) \underset{p\to\infty}{\sim} \frac{1}{\sigma\sqrt{2\pi p^3}}$.
- $P\left[\sup_{0\leq t\leq 1}\left|\frac{H_{[pt]}}{\sqrt{p}} - \frac{2}{\sigma^2}\frac{S_{[pt]}-I_{[pt]}}{\sqrt{p}}\right| > p^{-1/8} \,\middle|\, T_1 = p\right] < \exp\left(-p^{\varepsilon'}\right)$.
- $\left(H_k^{(p)}, 0 \leq k \leq p\right)$ a la même distribution que $(H_k, 0 \leq k \leq p)$ sous $P(\cdot \mid T_1 = p)$.
- Le théorème découle de l'inégalité précédente et du lemme suivant :

Lemme

La distribution du processus $\left(\frac{1}{\sigma\sqrt{p}} S_{[pt]}, 0 \leq t \leq 1\right)$ sous la probabilité conditionnelle $P(\cdot \mid T_1 = p)$ converge lorsque $p \to \infty$ vers la loi de l'excursion Brownienne normalisée.
TENSION!!

1. Notion d'arbre
2. Arbre de Galton Watson
3. Arbres réels

Arbre réel enraciné

Définition

Un espace métrique compact $(\mathcal{T}, d, \rho)$ est un **arbre réel** s'il vérifie les deux conditions suivantes : pour $a, b \in \mathcal{T}$

- Il existe une unique isométrie $f_{a,b}$ de $[0, d(a,b)]$ dans $\mathcal{T}$ tel que $f_{a,b}(0) = a$ et $f_{a,b}(d(a,b)) = b$.
- Si q est une application continue et injective de $[0,1]$ dans $\mathcal{T}$, tel $q(0) = a$ et $q(1) = b$, alors on a :
$$q([0,1]) = f_{a,b}([0, d(a,b)]).$$
- $\rho = \rho(\mathcal{T})$ est appelé **racine**, un élément de $\mathcal{T}$.

ascendants et descendants

Définition

- $[\![a,b]\!] := f_{a,b}([0,d(a,b)]$ est le **segment de l'arbre** reliant a à b.
- $[\![\rho,a]\!]$ est le **chemin reliant la racine à** a, et est interprété comme la ligne des ascendants de a.
- $a \preccurlyeq b$ (a est un **ascendant** de b) si et seulement si $a \in [\![\rho,b]\!]$.

Distance de Haussdorf

Pour étudier la convergence d'arbres réels, il nous faut une distance

Définition

Soit (E, δ) un espace métrique. La **distance de Hausdorff** est définie par :

$$\delta_{\text{Haus}}\left(K, K'\right) = \inf\left\{\varepsilon > 0 : K \subset U_{\varepsilon}\left(K'\right) \text{ and } K' \subset U_{\varepsilon}(K)\right\}$$

Distance de Gromov-Haussdorf

Définition

$\mathscr{T}$ et $\mathscr{T}'$ deux espaces métriques compacts de racines ρ et ρ'. On définit : $d_{GH}(\mathscr{T},\mathscr{T}')$ par :

$$d_{GH}(\mathscr{T},\mathscr{T}') = \inf\{\delta_{Haus}(\varphi(\mathscr{T}),\varphi'(\mathscr{T}')) \vee \delta(\varphi(\rho),\varphi'(\rho'))\},$$

où l'infininum est choisi sur tous les choix d'espaces métriques (E,δ) et toutes les isométries $\varphi : \mathscr{T} \longrightarrow E$ et $\varphi' : \mathscr{T}' \longrightarrow E$.

Arbres équivalents

Définition

Deux arbres réels enracinés $\mathscr{T}_1$ et $\mathscr{T}_2$ sont équivalents si et seulement s'il existe une isométrie bijective et racine-invariante de $\mathscr{T}_1$ sur $\mathscr{T}_2$.
$d_{GH}(\mathscr{T},\mathscr{T}')$ dépend seulement des classes d'équivalences de $\mathscr{T}$ et $\mathscr{T}'$.
On note $\mathbb{T}$ l'ensemble des classes d'équivalences d'arbres réels et on a que d_{GH} induit une distance sur $\mathbb{T}$.

Théorème

L'espace métrique $(\mathbb{T}, d_{GH})$ est complet et séparable.

Encodage

Soit $g : [0,\infty[\longrightarrow [0,\infty[$ à support compact tel que $g(0) = 0$. Pour $s, t \geq 0$, on définit :

$$m_g(s,t) = \inf_{r \in [s \wedge t, s \vee t]} g(r) \text{ et } d_g(s,t) = g(s) + g(t) - 2m_g(s,t)$$

- $s \sim t$ si et seulement si $d_g(s,t) = 0$ si et seulement si $g(s) = g(t)$.
- $\mathcal{T}_g = [0,\infty[/\sim$ et $d_g(s,t)$ induit une distance sur $\mathcal{T}_g$.
- Notons $p_g : \left[0,\infty\right[\longrightarrow \mathcal{T}_g$ la surjection canonique, continue.
- $\mathcal{T}_g = p_g([0,\zeta])$ est compact.

Théorème

L'espace métrique $(\mathcal{T}_g, d_g)$ est un arbre réel de racine $\rho = p_g(0)$.

Remarque : Tout arbre réel enraciné peut être codé par une fonction g et se représente de la forme $\mathcal{T}_g$.

Distortion

Définition

Soit $(\mathscr{T}_1, d_1)$ et $(\mathscr{T}_2, d_2)$ deux espaces métriques compact.
Une **correspondance** entre $\mathscr{T}_1$ and $\mathscr{T}_2$ est un sous-ensemble $\mathscr{R}$ de $\mathscr{T}_1 \times \mathscr{T}_2$ tel que :

- pour tout $x_1 \in \mathscr{T}_1$ il existe $x_2 \in \mathscr{T}_2$ tel que $(x_1, x_2) \in \mathscr{R}$
- pour tout $y_2 \in \mathscr{T}_2$ il existe $y_1 \in \mathscr{T}_1$ tel $\left(y_1, y_2\right) \in \mathscr{R}$.

La **distortion** de la correspondance $\mathscr{R}$ est définie par :

$$\mathrm{dis}(\mathscr{R}) = \sup\left\{\left|d_1\left(x_1, y_1\right) - d_2\left(x_2, y_2\right)\right| : (x_1, x_2), \left(y_1, y_2\right) \in \mathscr{R}\right\}$$

Distortion

Proposition

$$d_{GH}\left(\mathcal{T},\mathcal{T}'\right)=\frac{1}{2}\inf_{\mathscr{R}\in\mathscr{C}(\mathcal{T},\mathcal{T}'),(\rho,\rho')\in\mathscr{R}}\operatorname{dis}(\mathscr{R}).$$

Lemme

Soit g et f deux fonctions à support compact de $[0,\infty[$dans$[0,\infty[$ tel que $g(0)=f(0)=0$, alors :

$$d_{GH}\left(\mathcal{T}_g,\mathcal{T}_f\right)\leq 2\|g-f\|_\infty.$$

L'arbre continu aléatoire

Définition

L'arbre (**CRT**) est l'arbre réel $\mathscr{T}_{\mathbf{e}}$ codé par l'excursion brownienne normalisée.

- Le CRT $\mathscr{T}_{\mathbf{e}}$ est une variable aléatoire prenant ses valeurs dans l'ensemble $\mathbb{T}$

Convergence

Théorème

Pour tout $n \geq 1$, soit $\mathscr{T}_{(n)}$ un arbre fini aléatoire choisi uniformément dans $\mathbf{A}_n$. Alors $(2n)^{-1/2}\mathscr{T}_{(n)}$ converge en distribution vers le CRT $\mathscr{T}_{\mathbf{e}}$, dans l'espace $\mathbb{T}$.

Démonstration

- θ un arbre de Galton-Watson de distribution $\mu(k) = 2^{-k-1}$, $\sigma^2 = 2$
- θ_n choisi sous θ conditionné à avoir n branches.
- θ_n et $\mathscr{T}_{(n)}$ ont même distribution.
- $\left(C_t^n, t \geq 0\right)$ fonction contour de θ_n
- $\widetilde{C}_t^n = (2n)^{-1/2} C_{2nt}^n, \quad t \geq 0$
- $\left(\widetilde{C}_t^n, t \geq 0\right) \xrightarrow[p\to\infty]{(d)} (e_t, t \geq 0)$

Merci pour votre écoute !

Référence :

Random Trees and Applications, Jean-Francois LE GALL