

Elagage : des arbres de Galton-Watson aux arbres continus

Patrick Hoscheit

1 Rappels : Arbres de Galton-Watson

Nous allons d'abord décrire quels sont les arbres qui nous intéressent, les *arbres généalogiques* et plus particulièrement, dans un cadre aléatoire, les arbres de *Galton-Watson*. Une bonne référence pour tous ces sujets est le livre de Duquesne et Le Gall ([DLG02]) ainsi que l'article fondateur d'Aldous ([Ald91]).

Arbres discrets

Commençons par définir les objets en question, suivant Neveu ([Nev86]). On décrit les arbres de façon généalogique, par des suites d'entiers :

Définition 1. *Un arbre généalogique est une partie \mathcal{T} de l'espace $\mathcal{U} = \bigcup_{n=0}^{\infty} \mathbf{N}^n$ (où on note $\mathbf{N}^0 = \{\emptyset\}$ la racine), vérifiant les propriétés évidentes :*

- *Un arbre contient toujours sa racine : $\emptyset \in \mathcal{T}$*
- *Si un arbre contient un individu, il contient son parent : si $u = u_1 \dots u_n \in \mathcal{T}$, avec $n \geq 1$, alors $u_1 \dots u_{n-1} \in \mathcal{T}$.*
- *Chaque individu a un nombre fini d'enfants : pour tout $u \in \mathcal{T}$, il existe un entier $k_u(\mathcal{T}) \geq 0$ tel que $uj \in \mathcal{T}$ si et seulement si $j \in \llbracket 1, k_u(\mathcal{T}) \rrbracket$.*

Nous ne ferons pas la distinction, habituelle en théorie des graphes, entre noeuds et feuilles : tous les éléments d'un arbre seront appelés indifféremment *noeuds* ou *individus*. Les enfants de l'individu u sont exactement les individus notés $u1, u2, \dots, uk_u(\mathcal{T})$. On appellera *hauteur* (ou *génération*) de l'individu $u = u_1 \dots u_n$ l'entier n^1 et on appellera *taille* d'un arbre le nombre total $|\mathcal{T}|$ de ses individus.

Introduisons maintenant le codage d'arbres par des fonctions de contour. Rappelons tout d'abord ce qu'est le parcours d'un arbre en profondeur d'abord. Il s'agit de parcourir l'arbre, en partant de la racine \emptyset , et, arrivé en un noeud u , d'aller vers la premier enfant non visité uj ou alors, si tous les enfants ont été visités, de retourner en son parent. Le procédé s'arrête lorsqu'on est revenu à la racine en ayant visité tous les enfants de celle-ci.

1. Ainsi, la racine est de hauteur 0, ses enfants de hauteur 1, etc.

par récurrence par

$$Z_0 = 1 ; Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n+1}$$

Il faut se représenter $(Z_n, n \geq 0)$ comme modélisant l'évolution d'une population dont les individus ont chacun un nombre d'enfants suivant la loi μ , de façon indépendante les uns des autres. Notons $m = \sum k\mu(k)$ l'espérance de μ . Le comportement asymptotique des processus de Galton-Watson est décrit par le théorème suivant, qui découle essentiellement du fait que $(m^{-n}Z_n, n \geq 0)$ est une martingale pour la filtration canonique.

Théorème 3. *D'une part, on a $\mathbf{P}(\exists n, Z_n = 0) + \mathbf{P}(Z_n \rightarrow \infty) = 1$. D'autre part, on a les comportements suivants :*

- Si $m < 1$ (cas sous-critique), $m^{-n}Z_n$ tend p.s. vers 0. En particulier, $Z_n = 0$ pour n suffisamment grand.
- Si $m = 1$ (cas critique), $Z_n = 0$ pour n suffisamment grand.
- Si $m > 1$ (cas surcritique), $\mathbf{P}(\exists n, Z_n = 0) = \rho$, où ρ est l'unique point fixe de la fonction génératrice de μ dans $]0,1[$.

Dans le cas surcritique, la probabilité d'explosion est donc strictement positive. On montre en fait que $m^{-n}Z_n$ converge p.s. vers une variable W . Sur $\{W > 0\}$, la suite $(Z_n, n \geq 0)$ explose clairement de façon exponentielle. Une réciproque partielle a été apportée par Kesten et Stigum, qui ont montré que si $\sum k \ln(k)\mu(k) < \infty$, alors sur l'événement de non-extinction, $W > 0$. Nous supposons désormais que μ est de variance finie, donc que le processus surcritique explose exponentiellement avec probabilité positive.

Il est très facile de se représenter les processus de Galton-Watson sous forme d'arbres généalogiques, en reliant chaque individu à son parent et à ses descendants. De fait, on a le théorème suivant :

Théorème 4. *Il existe, pour toute loi de reproduction μ , une unique loi \mathbf{Q}_μ sur l'ensemble des arbres ordonnés enracinés, telle que $\mathbf{Q}_\mu(k_\emptyset = j) = \mu(j)$ et telle que, sous $\mathbf{Q}_\mu(\cdot | k_\emptyset = j)$, les j sous-arbres issus de \emptyset sont indépendants, de loi \mathbf{Q}_μ .*

Démonstration. On peut donner la forme explicite de cette loi si \mathcal{T} est un arbre fini :

$$\mathbf{Q}_\mu(\mathcal{T}) = \prod_{v \in \mathcal{T}} \mu(k_v) \tag{1}$$

On peut étendre cette loi aux arbres infinis par limite projective : la loi d'un arbre infini est donnée par la suite des lois de ses restrictions aux hauteurs $1, 2, \dots$, donc cette formule détermine bien une unique loi sur l'ensemble des arbres généalogiques². \square

2. Si l'on munit cet espace de la topologie produit évidente

2 Elagage d'arbres de Galton-Watson

Définition du procédé d'élagage

Nous allons exposer ici les travaux d'Aldous et Pitman dans ([AP98]), qui marquent la première étude de l'élagage comme définissant un processus à valeurs dans les arbres.

Définition 5. Soit \mathcal{T} un arbre généalogique aléatoire. Un élagage uniforme de \mathcal{T} est un processus aléatoire $(\mathcal{T}_u, 0 \leq u \leq 1)$, tel que $\mathcal{T}_1 = \mathcal{T}$ p.s. et tel que

$$(\mathcal{T}_u, 0 \leq u \leq 1) \stackrel{\mathcal{L}}{=} (\mathcal{T}(u), 0 \leq u \leq 1)$$

où on a construit la famille $(\mathcal{T}(u))$ de la façon suivante. Soit un arbre $\mathcal{T}(1)$ de même loi que \mathcal{T} et soit, sachant $\mathcal{T}(1)$, une famille (ξ_e) de v.a. iid uniformes sur $[0, 1]$ indexée par les arêtes de $\mathcal{T}(1)$. On note alors $\mathcal{T}(u)$ le sous-arbre généalogique contenant la racine et les arêtes telles que $\xi_e < u$.

Partant d'un arbre de Galton-Watson \mathcal{T} , dont la loi de reproduction est la loi de Poisson de paramètre r , pour $r > 0$, on peut ainsi définir un processus markovien réversible

$$(\mathcal{T}_u, 0 \leq u \leq 1)$$

avec $\mathcal{T}_1 = \mathcal{T}$ et qui soit élagué de façon *consistante*, c'est-à-dire que le processus $(\mathcal{T}_{ut}, 0 \leq u \leq 1)$ est un élagage uniforme de \mathcal{T}_t , pour tout $0 \leq t \leq 1$.

Dans le cas particulier poissonnien que nous avons décrit ici, on peut même faire mieux. En effet, on peut montrer facilement que la somme de N variables de Bernoulli de paramètre u , quand N suit une loi de Poisson de paramètre r , est encore une variable de Poisson, de paramètre ur . L'arbre élagué avec une intensité u est donc encore un arbre de Galton-Watson, de loi de reproduction poissonnienne de paramètre ur .

D'après la propriété de consistance de l'élagage, on peut ainsi définir³ un processus $(\mathcal{T}_\mu, \mu \geq 0)$, à trajectoires croissantes, tel que pour tout μ , l'arbre \mathcal{T}_μ soit un arbre de Galton-Watson, dont la loi de reproduction est la loi de Poisson de paramètre μ . On peut alors poser $\mu = \exp(-\theta)$, pour obtenir un processus (décroissant) indexé par $\theta \in \mathbf{R}$. Remarquons tout de suite que les arbres sont de nature différente suivant le signe de θ , comportement que l'on peut résumer par le diagramme de phase suivant :

On peut donner une description simple du processus forward, en décrivant explicitement les transitions entre \mathcal{T}_{θ_1} et \mathcal{T}_{θ_2} pour $\theta_2 < \theta_1$ deux réels.

Théorème 6. Soit $\theta_2 < \theta_1$. Conditionnellement à \mathcal{T}_{θ_1} , soit pour chaque $v \in \mathcal{T}_{\theta_1}$ une variable $N_{\theta_1, \theta_2}(v)$, de loi de Poisson de paramètre $\exp(-\theta_2) -$

3. Grâce au théorème de Kolmogorov

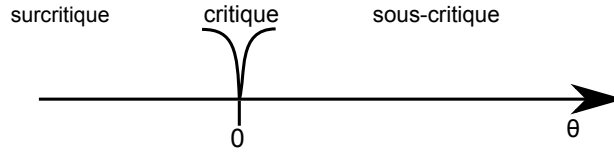


FIGURE 3 – Elagage d’un arbre de Galton-Watson surcritique

$\exp(-\theta_1)$ et, sachant les $N_{\theta_1, \theta_2}(v)$, soit $\hat{\mathcal{T}}_{\theta_2}$ obtenu à partir de \mathcal{T}_{θ_1} en attachant à chaque $v \in \mathcal{T}_{\theta_1}$, $N_{\theta_1, \theta_2}(v)$ copies indépendantes de \mathcal{T}_{θ_2} . Alors

$$(\mathcal{T}_{\theta_1}, \mathcal{T}_{\theta_2}) \stackrel{\mathcal{L}}{=} (\mathcal{T}_{\theta_1}, \hat{\mathcal{T}}_{\theta_2})$$

On sait que pour $\theta \geq 0$, l’arbre \mathcal{T}_θ est toujours fini p.s. Par contre, pour $\theta < 0$, l’arbre, de loi surcritique, est infini avec probabilité positive. On peut se poser la question du premier instant où cela arrive :

Définition 7. On appelle temps d’explosion le premier instant où l’arbre élagué devient infini, soit

$$A = \inf\{\theta \in \mathbf{R}, |\mathcal{T}_\theta| < \infty\}$$

Sachant que l’arbre \mathcal{T}_θ est un arbre de Galton-Watson, on peut écrire

$$\mathbf{P}(A \leq \theta) = \mathbf{P}(|\mathcal{T}_\theta| < \infty) = F(\theta)$$

Ainsi, la fonction de répartition de A est égale à la probabilité d’extinction $F(\theta)$. On sait par ailleurs (cf. le théorème 3) que $F(\theta)$ est la plus petite racine positive de l’équation

$$F(\theta) = \exp(-e^{-\theta}(1 - F(\theta))) \quad (2)$$

Notons $\bar{F}(\theta) = 1 - F(\theta)$ et posons, si $\theta \leq 0$, $\hat{\theta} = \theta - \log F(\theta)$. Remarquons que dans ce cas, $\hat{\theta} \geq 0$: on appelle $\hat{\theta}$ le *conjugué* de θ .

D’après la forme explicite de la loi des arbres \mathcal{T}_θ (équation 1), on voit qu’on a, pour $\theta \leq 0$,

$$\mathbf{P}(\mathcal{T}_\theta = \mathbf{t}) = F(\theta)\mathbf{P}(\mathcal{T}_{\hat{\theta}} = \mathbf{t}) \quad (3)$$

Ainsi, l’arbre surcritique \mathcal{T}_θ , conditionné à être fini, a la même loi que l’arbre sous-critique $\mathcal{T}_{\hat{\theta}}$. On peut maintenant facilement en déduire la loi de l’arbre juste avant l’explosion, conditionnellement à A :

Théorème 8. Pour $\theta \leq 0$, la loi conditionnelle de l’arbre T_{A-} sachant $\{A = \theta\}$ est donnée par

$$\mathbf{P}(T_{A-} = \mathbf{t} | A = \theta) = |\mathbf{t}|(1 - \exp(-\hat{\theta}))\mathbf{P}(T_{\hat{\theta}} = \mathbf{t})$$

Dans l'article original, Aldous et Pitman donnent en fait des résultats bien plus profonds, notamment en décrivant le processus d'élagage comme élagage d'un arbre infini : ils introduisent un arbre infini, constitué d'une colonne vertébrale infinie le long de laquelle s'accrochent des arbres de loi poissonnienne critique et considèrent un élagage uniforme $(T_\theta^*, \theta \geq 0)$ de cet arbre. Alors, pour $\theta < 0$, la loi de l'arbre T_θ^* est exactement la loi décrite par le terme de droite du théorème précédent. En d'autres termes, conditionnellement à A , l'arbre juste avant explosion a la loi de l'élagage d'un arbre infini : le biais s'interprète alors comme résultant de l'élagage de la colonne vertébrale infinie.

3 Extension aux arbres continus

La question se pose maintenant de l'extension de ces résultats élémentaires aux arbres continus. Nous allons donner d'abord une brève description des arbres continus browniens, introduits par Aldous ([Ald91],[Ald93]), avant de montrer comment on définit l'élagage sur ces arbres, en suivant les travaux d'Abraham, Delmas et Voisin ([ADV08],[AD09]).

Arbres continus browniens

Nous allons ici présenter l'arbre continu brownien comme limite d'échelle des arbres de Galton-Watson, bien que ce passage à la limite ne soit pas nécessaire en théorie. Commençons par regarder la limite des processus associés à l'arbre, le processus de Galton-Watson (appelé encore processus de branchement discret) qui représente la masse totale de l'arbre et le processus de contour, qui code l'arbre tout entier.

On sait depuis Lamperti ([Lam67]) que la seule limite d'échelle possible pour les processus de Galton-Watson est constituée par les processus de branchement à espace d'états continu (CSBP). Ce sont des processus à valeurs dans \mathbf{R}_+ , décrivant l'évolution d'une population soumise à un *branchement*. En particulier, ils sont caractérisés par une fonction ψ précisant les modalités du branchement. Dans le cas qui nous intéresse, donnons le théorème de convergence :

Théorème 9. *Soit μ une mesure de probabilité sur \mathbf{N} , telle que*

$$\sum k\mu(k) = 1; \quad 0 < \sum k^2\mu(k) = \sigma^2 < \infty$$

Considérons une suite $(N^k, k \geq 0)$ de processus de Galton-Watson de loi de reproduction μ . Alors, on a la convergence

$$\left(\frac{1}{k}N_{[kt]}^k, t \geq 0\right) \xrightarrow{k \rightarrow \infty} (Y_t, t \geq 0)$$

au sens de la convergence en loi des lois fini-dimensionnelles, où $(Y_t, t \geq 0)$ est une diffusion de Feller, qui est la solution de l'EDS

$$dY_t = \sigma \sqrt{Y_t} dB_t$$

avec $(B_t, t \geq 0)$ un mouvement brownien standard.

Dans la mesure où les processus de Galton-Watson codent pour la taille d'une population évoluant aléatoirement, on peut se poser la question de l'existence d'arbres qui décriraient de façon plus précise la généalogie d'une population décrite par une diffusion de Feller. Le théorème suivant est dû à Aldous ([Ald93]) :

Théorème 10. *Soit pour tout n , \mathcal{T}_n un arbre de Galton-Watson conditionné à une taille totale de population n , dont la loi de reproduction vérifie*

$$\sum k\mu(k) = 1; 0 < \sum k^2\mu(k) = \sigma^2 < \infty; \text{PGCD}(j, \mu(j) > 0) = 1$$

Alors, si $C_n : \llbracket 0, \zeta(n) \rrbracket \rightarrow \mathbf{R}_+$ est le processus de contour de l'arbre \mathcal{T}_n dont les branches sont de longueur $\sigma n^{-1/2}$, si \tilde{C}_n est le processus renormalisé pour s'annuler en 0 et en 1, soit $\tilde{C}_n(t) = C_n(\zeta(n)t)$ avec interpolation linéaire, on a la convergence en loi

$$(\tilde{C}_n(t), 0 \leq t \leq 1) \longrightarrow (2B_t, 0 \leq t \leq 1)$$

au sens de la topologie de Skorokhod sur $\mathcal{C}([0, 1])$, et où B est une excursion brownienne normalisée standard.

On peut donc voir l'excursion brownienne comme codant la généalogie d'un arbre. La figure suivante montre comment on peut lire la généalogie de quatre individus : si $t \in [0, 1]$, la génération de l'individu t est donnée par B_t et la génération du plus récent ancêtre commun entre deux individus $t \leq t'$ est donnée par $\inf_{t \leq s \leq t'} B_s$. La figure représente, si l'on veut, une marginale de l'arbre complet, qui correspond à cette construction pour tous les individus (la limite projective). En particulier, on peut voir sur cette figure l'une des caractéristiques essentielles de l'arbre codé par l'excursion brownienne : tous ses noeuds sont de degré 3. Ceci est dû au fait que les minima locaux du mouvement brownien sont toujours distincts p.s.

Définition 11. *Soit $(B_t, 0 \leq t \leq \sigma)$ une excursion brownienne, codant pour un arbre brownien continu ρ . On appelle masse de l'arbre ρ le réel σ , strictement positif.*

Ces résultats de convergence browniens ont été généralisés par Le Gall et Le Jan ([LGLJ98]) au cas d'un branchement plus général que le branchement binaire associé à l'excursion brownienne, décrit par une fonction ψ .

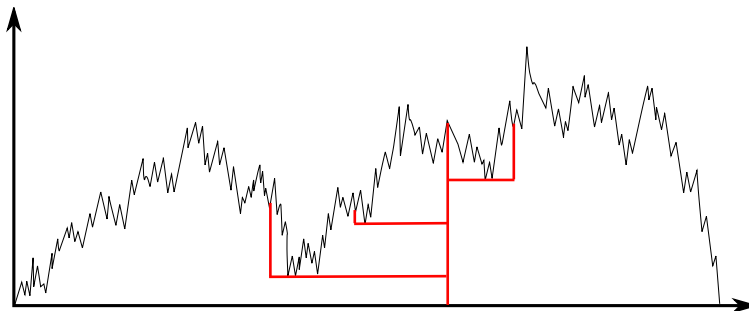


FIGURE 4 – L’arbre des ancêtres communs de quatre individus

Le tableau est donc le suivant : aux processus de Galton-Watson discrets correspondent des processus de branchement continu (CSBP) ; aux arbres de Galton-Watson (sous)critiques, décrivant la généalogie des processus correspondent des arbres de Lévy (sous)critiques ; pour décrire ces arbres, on utilise des fonctions de contour définies sur $[[0, \zeta(\mathcal{T})]]$ pour les arbres de Galton-Watson, des fonctions de hauteur, définies sur $[0, \sigma]$ pour les arbres de Lévy.

Dans la suite, nous nous focaliserons sur le cas des arbres continus browniens, cas particuliers des arbres de Lévy quand le CSBP est un mouvement brownien réfléchi et quand le processus de hauteur est une excursion brownienne.

Elagage d’arbres browniens

On ne peut pas généraliser directement la percolation au cas des arbres continus. En effet, si on se donne un individu pris au hasard, de génération H_t , on peut représenter graphiquement son ascendance et les branchements qui ont donné naissance à cet individu, puisque les ancêtres de l’individu à la génération H_t sont exactement les individus que la « branche » $\{t\} \times [0, H_t]$ peut « voir ». En particulier, on sait que l’ancêtre commun aux individus t et t' est le minimum $\min_{s \in [t, t']} B_t$. Réciproquement, à tous les minima locaux du mouvement brownien réfléchi B_t correspond un ancêtre commun à t et à un autre individu, autrement dit un branchement de l’arbre. De la forme des trajectoires du mouvement brownien, on déduit alors que sur la « branche » $\{t\} \times [0, H_t]$, il y a une infinité de noeuds correspondant à des branchements.

Il est donc impossible de définir la percolation comme dans le cas discret, sous peine de n’avoir que des arbres triviaux. L’idée est alors d’exploiter la structure continue de l’arbre et de couper les branches grâce à un processus de Poisson. Ainsi, sur chaque « branche » $\{t\} \times [0, H_t]$, on dispose des marques, de telle sorte que, dans un « intervalle » de longueur ℓ , le nombre de marques suive une loi de Poisson de paramètre $\theta\ell$, où $\theta > 0$.

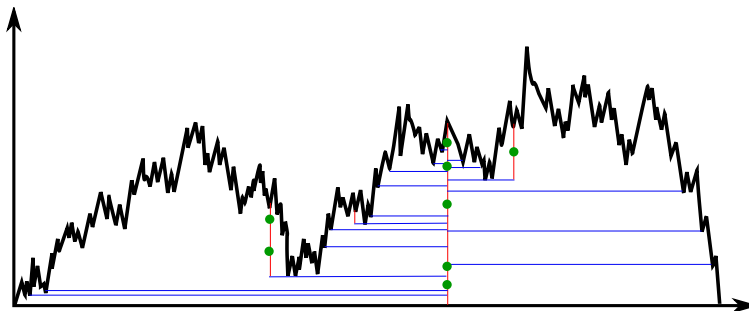


FIGURE 5 – Un morceau de CRT avec des marques d'élagage

Il n'est pas évident qu'un tel marquage puisse être défini correctement : en effet, il faut que le marquage soit consistant, car deux individus pris au hasard ont une ascendance commune jusqu'à un certain point. Le marquage doit donc être défini de telle sorte que les marques jusqu'au plus récent ancêtre commun de ces individus soient les mêmes. Cette difficulté contournée (construction dite du *serpent poissonnier*), on peut ainsi construire, pour $\theta > 0$ fixé, un arbre ρ^θ , qui est le sous-arbre contenant la racine une fois les branches coupées aux endroits marqués.

En fait, Abraham et Delmas construisent quelque chose de plus fort, puisqu'ils parviennent à transposer dans le cas continu l'idée de couplage du cas discret : ils arrivent à construire un processus $(\rho^\theta, \theta \geq 0)$, dont les trajectoires correspondent à l'élagage de plus en plus intense du même arbre, c'est-à-dire dont les trajectoires sont décroissantes au sens des arbres.

Théorème 12. *Il existe un processus aléatoire $(\rho^\theta, \theta \geq 0)$ à valeurs dans les arbres continus, tel que θ^0 a la loi d'un CRT brownien et tel que p.s., pour tout $\theta' > \theta$, l'arbre $\rho^{\theta'}$ est obtenu à partir de ρ^θ par élagage. Ainsi, le processus $(\rho^\theta, \theta \geq 0)$ est markovien.*

La question se pose alors, comme dans le cas discret, de la loi de l'arbre élagué et, éventuellement l'existence d'une limite projective, processus à valeurs arbres qui engloberait toute une famille de CRT. Une première obstruction est liée au fait qu'il n'existe pas, a priori, de description de l'arbre continu brownien *surcritique*⁴. Abraham et Delmas définissent l'arbre surcritique en remarquant qu'en coupant un arbre sous-critique à une certaine hauteur a et en effectuant un changement de mesure absolument continu par rapport à la loi de cet arbre coupé, la mesure obtenue peut être vue comme décrivant un arbre surcritique coupé au niveau a . Ainsi, par consistance, on peut définir la loi d'un arbre surcritique. On peut donc considérer le processus d'élagage $(\rho^\theta, \theta \in \mathbf{R})$ limite projective des processus précédents,

4. on ne peut pas décrire l'arbre surcritique à partir d'excursions browniennes, puisque celles-ci sont de longueur finie et que l'arbre surcritique est intuitivement de masse finie

tels que pour tout $\theta \in \mathbf{R}$, $(\rho^{\theta+\gamma}, \gamma \geq 0)$ soit un élagage de ρ^θ . Autrement dit, $(\rho^\theta, \theta \in \mathbf{R})$ est un processus décroissant à valeurs arbres qui réalise en quelque sorte l'analogue continu du processus $(\mathcal{T}_\theta, \theta \in \mathbf{R})$ d'Aldous et Pitman.

On démontre alors les théorèmes suivants, que l'on pourra comparer aux analogues discrets :

Théorème 13. *Sous la mesure d'excursion du mouvement brownien \mathbf{N} , la loi du temps d'explosion $A = \inf\{\theta \in \mathbf{R}, \sigma_\theta < \infty\}$ admet pour densité 2λ où λ est la mesure de Lebesgue sur $(-\infty, 0)$.*

Théorème 14. *Soit $\theta < 0$. Alors, conditionnellement à $\{A = \theta\}$, l'arbre juste avant explosion ρ_A a la même loi qu'un arbre critique conditionné à être infini, élagué avec une intensité $|\theta|$.*

Nous ne donnerons pas de définition précise de cet arbre critique conditionné à être infini, mais on peut le voir, de la même façon que l'arbre qu'Aldous et Pitman considèrent dans leur article, comme une colonne vertébrale infinie sur laquelle se greffent des arbres finis, de loi critique.

Références

- [AD09] R. Abraham and J.F. Delmas. A continuum-tree-valued Markov process. *Arxiv preprint arXiv :0904.4175*, 2009.
- [ADV08] R. Abraham, J.F. Delmas, and G. Voisin. Pruning a Lévy continuum random tree. *Arxiv preprint arXiv :0804.1027*, 2008.
- [Ald91] D. Aldous. The continuum random tree. I. *The Annals of Probability*, 19(1) :1–28, 1991.
- [Ald93] D. Aldous. The continuum random tree III. *The Annals of Probability*, pages 248–289, 1993.
- [AP98] D. Aldous and J. Pitman. Tree-valued markov chains derived from galton-watson processes. *Annales de l'IHP / Probabilités et statistiques*, 34(5) :637–686, 1998.
- [DLG02] T. Duquesne and J.F. Le Gall. Random trees, Levy processes and spatial branching processes, volume 281. *Asterisque*, 2002.
- [Lam67] J. Lamperti. The limit of a sequence of branching processes. *Probability Theory and Related Fields*, 7(4) :271–288, 1967.
- [LGLJ98] J.F. Le Gall and Y. Le Jan. Branching processes in Levy processes : The exploration process. *The Annals of Probability*, 26(1) :213–252, 1998.
- [Nev86] J. Neveu. Arbres et processus de Galton-Watson. *Annales de l'IHP / Probabilités et statistiques*, 22(2) :199–207, 1986.