Mutations sur un arbre aléatoire binaire mesuré

Jean-Jil DUCHAMPS
sous la direction d'Amaury LAMBERT

Juin 2016

Table des matières

Introduction

En génétique des populations, on représente une population par un arbre, chaque feuille correspondant à un individu, et les lignées se séparant au niveau des branchements de l'arbre. On s'intéresse aux différentes mutations que peuvent subir un ou plusieurs gènes particuliers au cours de l'évolution, mutations que l'on peut représenter par des marques le long des branches de l'arbre. La population est partitionnée en différents groupes en fonction des allèles portés, et la loi de cette répartition intéresse les biologistes qui peuvent comparer des statistiques réelles aux répartitions théoriques des différents modèles. Par exemple, pour un modèle d'évolution classique, l'arbre de Kingman, pourvu de mutations sur les branches selon un taux constant au cours du temps, la loi de la partition allélique est donnée par la célèbre "*Ewens' Sampling Formula*" [4]. Pour un modèle d'arbre différent, les *splitting trees*, Nicolas Champagnat et Amaury Lambert [3] parviennent à caractériser en moyenne cette répartition.

Le but de ce mémoire est d'étendre cette dernière approche à un modèle d'arbre aléatoire particulier : les *coalescent point processes*, munis de mutations disposées le long des branches selon un processus Poisonnien. On considère un arbre aléatoire infini dont la frontière (qui représente la population échantillonnée au temps présent) est mesurée par une mesure aléatoire finie. La partition allélique est donc composée de blocs qui comprennent une infinité d'individus mais auxquels on peut assigner une masse finie. On étudie alors plusieurs propriétés de ce modèle : la loi de l'arbre des clones de l'origine, la loi de la suite des tailles des blocs de la partition allélique (spectre de fréquences alléliques). L'intérêt mathématique de cette étude est que l'on manipule des objets généraux et variés (arbres réels, processus ponctuels de Poisson, subordinateurs) qui ont de nombreuses applications dans la théorie des processus stochastiques.

On définit dans la section 1 les arbres aléatoires étudiés. Dans la section 2 on construit les mutations le long des branches de l'arbre, et on fait le lien entre différentes descriptions du modèle. Dans la section 3 on explicite un processus Markovien à valeurs dans les arbres réels qui décrit l'évolution de la sous-population clonale de l'arbre quand l'on fait varier le taux de mutation. Dans la 4e et dernière section on étudie diverses statistiques de la population clonale pour en déduire des résultats sur l'intensité du spectre de fréquences allélique.

1 Préliminaires

1.1 Arbres discrets, arbres réels

Un arbre, en théorie des graphes, est un graphe connexe sans cycle. On appellera ce genre de graphes les arbres discrets, et on considérera leurs sommets étiquetés selon la notation de Neveu par $\mathcal{U}$, l'ensemble des mots finis dans l'alphabet des nombres entiers :

$$\mathcal{U} = \bigcup_{n\geq 0} \mathbb{N}^n = \{u_1u_2\dots u_n,\ \ u_i \in \mathbb{N},\, n \geq 0\},$$

avec par convention $\mathbb{N}^0 = \{\emptyset\}$.

Définition 1. Un **arbre discret enraciné** est un sous-ensemble $\mathcal{T}$ de $\mathcal{U}$ tel que

1. $\emptyset \in \mathcal{T}$.
2. Pour $u = u_1 \dots u_n \in \mathcal{T}$, pour $1 \leq k < n$, on a $u_1 \dots u_k \in \mathcal{T}$.
3. Pour $u \in \mathcal{T}$ et $i \in \mathbb{N}$ tel que $ui \in \mathcal{T}$, pour $1 \leq j < i$, on a $uj \in \mathcal{T}$.

Pour $u, v \in \mathcal{T}$, s'il existe $w \in \mathcal{U}$ tel que $v = uw$, alors on dit que u est **ancêtre** de v et on note $u \prec v$. Dans le cas général, on note $u \wedge v$ le dernier ancêtre commun à u et v, c'est-à-dire le plus long mot $u_0 \in \mathcal{T}$ tel que $u_0 \prec u$ et $u_0 \prec v$.

Les arêtes de $\mathcal{T}$ en tant que graphe sont placées entre les parents u et leurs enfants ui (voir figure 1).

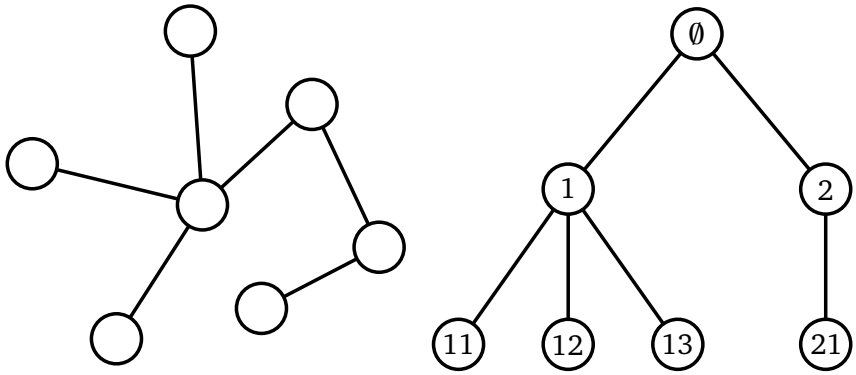

FIGURE 1 – Arbre comme graphe et une de ses représentations valables comme arbre discret

Si l'on assigne une longueur à chaque arête d'un arbre discret, on obtient ce que l'on appelle un arbre réel, que l'on peut voir comme un recollement d'intervalles de $\mathbb{R}$. Plus généralement, on peut définir les arbres réels comme les espaces métriques qui vérifient certaines propriétés.

Définition 2. Un espace métrique $(\mathbb{T}, d)$ est un **arbre réel** si pour tous $x, y \in \mathbb{T}$,

— il existe une unique isométrie $f_{x,y} : [0, d(x, y)] \to \mathbb{T}$ telle que $f_{x,y}(0) = x$ et $f_{x,y}(d(x, y)) = y$,

— tous les chemins continus injectifs de x à y ont la même image (le chemin $f_{x,y}([0, d(x, y)])$ de x à y dans $\mathbb{T}$).

Cet unique chemin géodésique de x à y est noté $[\![x, y]\!]$. Le **degré** d'un point $x \in \mathbb{T}$ est définit comme le nombre de composantes connexes de $\mathbb{T} \setminus \{x\}$, de sorte que l'on puisse définir :

— Les **feuilles** de $\mathbb{T}$ sont les points de degré 1.

— Les **nœuds internes** de $\mathbb{T}$ sont les points de degré 2.

— Les **points de branchements** de $\mathbb{T}$ sont les points de degré supérieur ou égal à 3.

On peut enraciner un arbre réel en distinguant un point $\rho \in \mathbb{T}$, appelé la **racine**. Un arbre réel enraciné est dit **simple** s'il peut être défini à partir d'un arbre discret en assignant une longueur à chaque arête.

À partir de cette définition, il est facile de voir que pour un arbre enraciné $(\mathbb{T}, d, \rho)$, pour tous $x, y \in \mathbb{T}$, il existe un unique point $a \in \mathbb{T}$ tel que $[\![\rho, x]\!] \cap [\![\rho, y]\!] = [\![\rho, a]\!]$.

On appelle a le *plus récent ancêtre commun* de x et y, et on le note $x \wedge y$. On a aussi une relation d'ordre intrinsèque à l'arbre enraciné : si $x \wedge y = x$, c'est-à-dire si $x \in [\![\rho, y]\!]$, on dit que x est un ancêtre de y, et on note $x \prec y$. On peut aussi vérifier une propriété des arbres réels, qui nous sera utile.

Proposition 1.1 (Condition des quatre points)**.** *Un arbre réel vérifie la* condition des quatre points *: pour $x, y, z, t \in \mathbb{T}$, on a*

$$d(x, y) + d(z, t) \leq \max(d(x, z) + d(y, t), d(x, t) + d(y, z)). \tag{1}$$

Cette condition veut dire qu'étant donnés quatre points, on se retrouve toujours dans une situation similaire à celle illustrée sur la figure 2, où l'on a

$$d(x, y) + d(z, t) \text{ (en bleu)} \ < d(x, z) + d(y, t) = d(x, t) + d(y, z)) \text{ (en rouge)} \ .$$

Démonstration. Soit $x, y, z, t \in \mathbb{T}$. Dans le sous-arbre de $\mathbb{T}$ engendré par ces quatre points, x est une feuille au bout de sa branche $[\![u_x, x]\!] := [\![y, x]\!] \cap [\![z, x]\!] \cap [\![t, x]\!]$ (remarquons que si par exemple $x \in [\![y, z]\!]$, alors techniquement x n'est pas une feuille, et $u_x = x$, mais le raisonnement suivant tient toujours). On peut donc remplacer x par tout point $x' \in [\![u_x, x]\!]$ sans changer la différence entre le terme à droite et le terme à gauche dans (1). On remplace donc x par u_x, et y, z, t sont remplacés de la même manière. Nécessairement, aucun des points restants n'est tout seul (c'est-à-dire $x \in \{y, z, t\}$), donc on a $x = y$ et $z = t$, ou bien $x = z$ et $y = t$, ou bien $x = t$ et $y = z$, et on vérifie facilement la condition. □

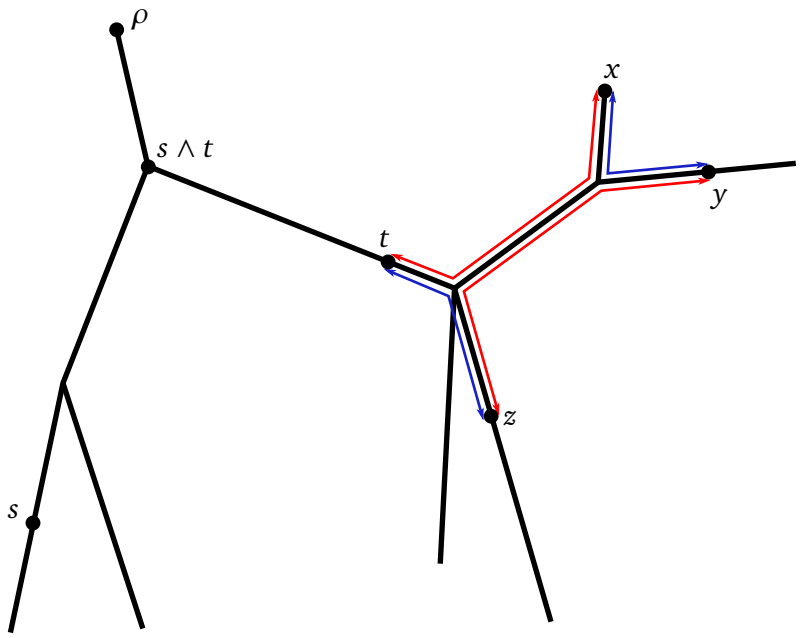

FIGURE 2 – Arbre réel (simple) enraciné

On se restreindra au cas des arbres simples, c'est-à-dire qui peuvent être construits en assignant des longueurs aux branches d'un arbre discret.

Définition 3. Un **arbre réel simple** est la donnée de $(\mathcal{T}, (\alpha(u), \zeta(u), \omega(u))_{u\in\mathcal{T}})$, où $\mathcal{T} \subset \mathcal{U}$ est un arbre discret enraciné, et $\alpha(u)$, $\zeta(u)$, $\omega(u)$ sont des réels appelés le temps de naissance, la durée de vie et le temps de mort de u, qui vérifient :

$$\zeta(u) = \omega(u) - \alpha(u) > 0,$$

$$\forall u \in \mathcal{T}, \forall i \in \mathbb{N}, \quad ui \in \mathcal{T} \Longrightarrow \alpha(ui) = \omega(u).$$

On peut vérifier qu'un tel objet est bien associé à un arbre réel enraciné $(\mathbb{T}, d, \rho)$, avec

$$\rho := (\emptyset, \alpha(\emptyset)),$$

$$\mathbb{T} := \{\rho\} \cup \bigcup_{u\in\mathcal{T}} \{u\}\times]\alpha(u), \omega(u)] \quad \subset \mathcal{U} \times \mathbb{R},$$

$$d((u, x), (v, y)) := \begin{cases} |x - y| & \text{si } u \prec v \text{ ou } v \prec u, \\ x + y - 2\omega(u \wedge v) & \text{sinon.} \end{cases}$$

Cela définit bien un arbre réel, et on remarque que $(u, x) \wedge (v, y) = (u \wedge v, \omega(u \wedge v))$.

Dans la suite, on construit des arbres réels simples aléatoires à partir d'objets connus, notamment les mesures ponctuelles de Poisson. Ces arbres seront en plus munis de marques le long des branches. On pourra voir les arbres comme des arbres phylogénétiques et les marques comme des mutations qui apparaissent au cours de l'évolution. Le but est d'étudier ce genre d'objets, en particulier la population clonale de l'arbre (les individus qui ne portent pas de mutations, sous-arbre noir de la figure 3), et sa structure quand on fait varier le taux de mutation.

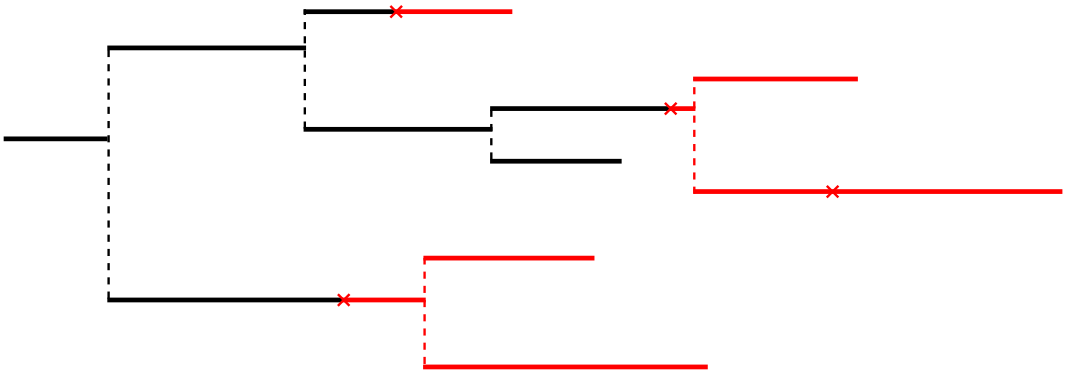

FIGURE 3 – Arbre simple avec mutations

1.2 Espace en peigne, CPP

Fixons un arbre réel enraciné $(\mathbb{T}, d, \rho)$. On considère la sphère de centre ρ et de rayon $a > 0$,

$$\mathbb{T}^{\{a\}} := \{x \in \mathbb{T}, d(x, \rho) = a\},$$

munie de la distance induite par d. Comme $\mathbb{T}$ est un arbre réel, il vérifie la condition des quatre points (proposition 1.1), ce qui implique, pour $x, y, z \in \mathbb{T}^{\{a\}}$:

$$d(x, y) + d(z, \rho) \leq \max(d(x, z) + d(y, \rho), d(y, z) + d(x, \rho)).$$

Or par définition, $d(x, \rho) = d(y, \rho) = d(z, \rho) = a$, on peut donc simplifier :

$$d(x, y) \leq \max(d(x, z), d(y, z)).$$

On dit alors que $(\mathbb{T}^{\{a\}}, d)$ est un espace ultramétrique. En fait, tout espace ultramétrique compact peut être vu comme la sphère d'un arbre. En effet, il existe toujours une représentation par une fonction en peigne de ces espaces, et ces représentations mettent en évidence une structure d'arbre.

Fonction en peigne Soit une fonction f d'un intervalle compact $I \subset \mathbb{R}$ dans $\mathbb{R}_+$ qui vérifie que pour tout $\epsilon > 0$, l'ensemble $\{f > \epsilon\}$ est fini. C'est-à-dire que f s'écrit $f = \sum_n a_n \mathbf{1}_{\{x_n\}}$, avec (x_n) suite injective à valeurs dans I, et (a_n) suite de réels positifs tels que $a_n \to 0$ quand $n \to \infty$. On appelle un tel couple (f, I) une fonction en peigne. On définit alors la pseudo-distance ultramétrique

$$d_f(x, y) = 2 \max_{]x,y]} f$$

sur l'espace I. On peut ensuite poser $\overset{\circ}{I} = I/\sim$ le quotient de cet espace par la relation d'équivalence définie par $x \sim y$ si et seulement si $d_f(x, y) = 0$. On vérifie facilement alors que $(\overset{\circ}{I}, d_f)$ est un espace ultramétrique, et que son complété $\bar{I}$ est un espace ultramétrique compact. Remarquons que I se projette canoniquement sur $\overset{\circ}{I}$, et que $\overset{\circ}{I}$ s'injecte canoniquement dans $\bar{I}$. De plus on peut vérifier que ces applications sont mesurables pour les Boréliens associés à la distance usuelle (resp. d_f) sur I (resp. $\overset{\circ}{I}$ et $\bar{I}$).

$$I \xrightarrow{p} \overset{\circ}{I} \xrightarrow{i} \bar{I}.$$

On peut donc transporter la mesure de Lebesgue sur I par l'application $i \circ p$, ce qui nous donne une mesure Leb $\circ\, (i \circ p)^{-1}$ sur $\bar{I}$, que l'on notera Leb par un léger abus de langage. Il existe une réciproque : tout espace ultramétrique compact (mesuré) est isomorphe à l'espace engendré par une fonction en peigne f, si l'on choisit f correctement [5].

En fait, on peut directement dessiner un arbre réel, à partir de la fonction f, tel que l'espace ultramétrique se retrouve comme la sphère de l'arbre. Pour dessiner l'arbre (voir figure 4), il suffit de :

— tracer des traits verticaux sous les points de f, c'est-à-dire tous les segments $[(t, 0), (t, f(t))]$, pour tout $t \in I$ tel que $f(t) > 0$. On trace aussi un segment suffisamment haut à l'extrémité gauche de I (que l'on appellera branche origine). Ces segments constituent le squelette de l'arbre.

— relier chaque point $(t, f(t))$ par un trait horizontal vers la gauche, jusqu'au premier segment vertical déjà tracé. Ces traits horizontaux ne font pas partie de l'arbre, ils indiquent un branchement (on identifie les deux extrémités de chaque segment horizontal).

Plus formellement, pour $z > \max_I f$ et t_0 l'extrémité gauche de I, on définit la racine ρ et le squelette Sk de l'arbre par :

$$\rho := (t_0, z)$$

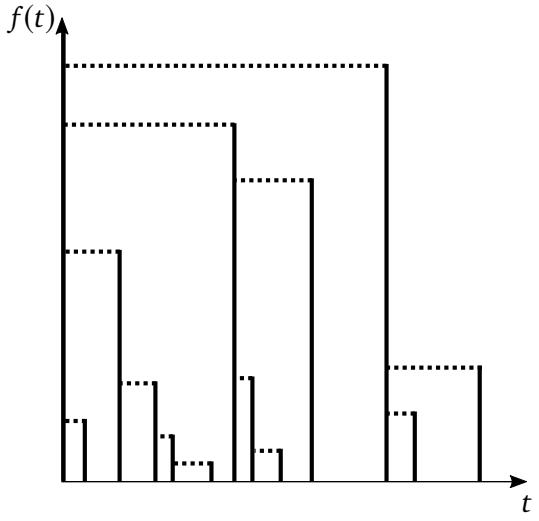

FIGURE 4 – Fonction en peigne, arbre associé

$$\mathrm{Sk} := \{t_0\}\times]0,\, z] \cup \{(t,\, y) \in I\times]0,\, z[,\, f(t) > y\},$$

et on définit la distance d_f sur ce squelette de manière cohérente avec la distance sur I : pour $t \leq s$, avec $(t,\, x)$, $(s,\, y) \in \mathrm{Sk}$,

$$d_f((t,\, x),\, (s,\, y)) = \begin{cases} |\max_{]t,s]} f - x| + |\max_{]t,s]} f - y| & \text{si } t < s, \\ |x - y| & \text{si } t = s. \end{cases}$$

On peut bien vérifier que cette définition fait de Sk un arbre réel, et les conditions sur la fonction en peigne f assurent que c'est même un arbre réel simple. Sa complétion, c'est-à-dire lorsqu'on rajoute les feuilles (l'espace $\bar{I}$, muni de la mesure Leb) est l'arbre réel défini par la fonction en peigne f.

C'est cet arbre, dit ultramétrique (toutes les feuilles sont à la même distance de la racine), qui nous intéresse. On peut le rendre aléatoire en rendant f aléatoire, par exemple de la façon suivante. Soit ν une mesure sur $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{\infty\}$, telle que pour tout $\epsilon > 0$, $\overline{\nu}(\epsilon) := \nu([\epsilon, \infty]) < \infty$. On pose $\mathcal{N}$ une mesure ponctuelle de Poisson sur $\mathbb{R}_+^2$ d'intensité $\mathrm{d}t \otimes \nu$. On définit alors f comme la fonction dont $\mathcal{N}$ est le graphe (privé des points de $\mathbb{R}_+ \times \{0\}$).

$$f^{\mathcal{N}}(t) = \begin{cases} x & \text{si } (t,\, x) \in \mathcal{N}, \\ 0 & \text{sinon.} \end{cases}$$

Avec $z > 0$ fixé tel que $\overline{\nu}(z) > 0$, on pose $T(z) := \inf\{t \geq 0,\, f^{\mathcal{N}}(t) \geq z\}$. On peut maintenant définir notre arbre aléatoire.

Définition 4. L'arbre ultramétrique aléatoire engendré par $I = [0,\, T(z)]$ et $f^{\mathcal{N}}_{|I}$ est appelé le **processus de points coalescents (CPP)** d'intensité ν et de hauteur z, noté $\mathrm{CPP}(\nu,\, z)$. On le munit implicitement de la mesure Leb, supportée par les feuilles, qui est simplement l'image de la mesure de Lebesgue sur l'intervalle $[0,\, T(z)]$.

1.3 Processus de naissance et de mort

Un processus de naissance et de mort est un processus qui compte une population qui évolue au cours du temps. On s'intéressera à la généalogie de la population elle-même, vue comme un arbre aléatoire, ce qui servira de description alternative pour nos arbres engendrés par des CPP.

On considère une population indexée par $\mathcal{U} = \bigcup_n \mathbb{N}^n$ qui évolue dans le temps de la manière suivante.

— Au temps $t_0 \in \mathbb{R}$, la population compte un seul individu $\emptyset$ (l'individu racine).
— Chaque individu u vivant au temps $t \geq t_0$, indépendamment de tous les autres, peut brancher (mourir et se reproduire) à taux $b(t)$. C'est-à-dire que l'instant de sa mort $\omega(u)$ est le premier atome d'une mesure de Poisson ponctuelle d'intensité $\mathbf{1}_{s \geq t} b(s)\mathrm{d}s$.
— À l'instant t de sa mort, il donne naissance à $N_t(u)$ nouveaux individus $u1$, $u2$, …, $uN_t(u)$. On remarque que si $N_t(u) = 0$, alors la lignée s'éteint effectivement.

De telles règles permettent clairement de définir un arbre réel simple, en associant les temps de naissance et de mort à tous les individus qui apparaissent au cours de l'évolution du processus.

Exemple L'arbre de Yule est un tel arbre binaire (c'est-à-dire $N \equiv 2$), avec un taux de branchement $b(t)$ constant égal à 1, c'est-à-dire que les branches qui séparent deux points de branchement sont des variables *i.i.d* exponentielles de paramètre 1.

2 Mutations sur un CPP

Il est classique d'étudier des mutations qui apparaissent sur des arbres aléatoires. On considère que les individus au temps présent sont les feuilles de l'arbre, et les mutations qui arrivent le long des branches sont uniques (deux mutations à deux endroits de l'arbre différents sont différentes). Chaque individu porte l'information de toutes les mutations sur sa lignée, c'est ce qui le caractérise. On appelle l'ensemble des mutations que porte un individu son haplotype, ou allèle. Si l'on regroupe les individus de la population finale par allèle, on voudrait connaître la répartition la population dans ces différentes classes. Par exemple, on peut se demander le nombre d'allèles différents qui apparaissent dans l'arbre, ou le nombre d'individus portant le même allèle fixé. Un objet fréquemment étudié est le spectre de fréquence allélique $(A_k)_{k \geq 1}$, où A_k est le nombre d'allèles qui sont portés par exactement k individus. On sait décrire des statistiques de ce genre dans différents cas, par exemple pour le coalescent de Kingman avec la célebrée *Ewens' Sampling Formula* [4] ou dans le cas des *splitting trees* [3]. Ici on essaye de construire des statistiques analogues dans le cas où l'on observe une population infinie (ν est de masse infinie), que l'on peut mesurer (avec la mesure de Lebesgue sur l'axe des abscisses). Précisons d'abord l'arbre aléatoire étudié, et la façon de placer les mutations.

2.1 Construction

Soit ν une mesure sur $\mathbb{R}_+ \cup \{\infty\}$, et μ une mesure sur $\mathbb{R}_+$. On fait les hypothèses suivantes :

$$\begin{gathered} \forall x > 0, \qquad \overline{\nu}(x) := \nu([x, \infty]) < \infty \ \text{ et } \ \underline{\mu}(x) := \mu([0, x]) < \infty, \\ \underline{\mu}(\infty) = \infty, \\ \nu \text{ et } \mu \text{ n'ont pas d'atomes sur } \mathbb{R}_+. \end{gathered} \tag{H}$$

Considérons le CPP d'intensité ν et de hauteur $z > 0$, que l'on veut marquer le long des branches au taux μ.

On rappelle que cet arbre est construit à partir d'une mesure ponctuelle de Poisson $\mathcal{N}$ d'intensité $dt \otimes \nu$ sur $\mathbb{R}_+^2$, et l'on dispose d'une racine ρ identifiée au point $(0, z)$. Pour jeter des mutations le long des branches, on définit indépendamment pour chaque point $N = (t, x)$ de $\mathcal{N} \cup \{(0, z)\}$, une mesure ponctuelle de Poisson M_N d'intensité μ sur $[0, x]$. Ainsi on définit bien une mesure ponctuelle de Poisson M le long du squelette de l'arbre. On a en particulier, pour chaque point $t \in \mathbb{R}_+$ de l'axe des abscisses (identifié aux feuilles de l'arbre), l'ensemble des mutations autres que celles de la branche origine sur la lignée issue de t est

$$M_t := \sum_{i \geq 1} M_{N_i}(\cdot \cap]x_{i+1}, x_i]),$$

où l'on définit $N_i = (t_i, x_i)$ le i-ième point de $\mathcal{N}$ dans la lignée de t. C'est-à-dire

$$x_1 = \max\{x \in \mathbb{R}_+, (s, x) \in \mathcal{N}, 0 < s \leq t\},$$

$$x_{i+1} = \max\{x \in \mathbb{R}_+, (s, x) \in \mathcal{N}, t_i < s \leq t\}.$$

M_t ainsi définie est une mesure ponctuelle de Poisson d'intensité μ sur $[0, x_1]$. Ces mutations sont bien compatibles avec la structure d'arbre : pour $t < s \in \mathbb{R}_+$ qui coalescent à la hauteur $x > 0$, on a $(M_t)_{|]x,\infty[} = (M_s)_{|]x,\infty[}$, et $(M_t)_{|[0,x]}$ indépendant de $(M_s)_{|[0,x]}$.

On a donc construit $\mathbb{T}^z$ un arbre CPP(ν, z), et M une mesure ponctuelle aléatoire sur $\mathbb{T}^z$. À partir de cette construction, on peut définir l'**arbre clonal** A_μ^z comme le sous-arbre du CPP qui est constitué des points x tels que la branche $[\![\rho, x]\!]$ qui les relie à la racine ne porte pas de mutation :

$$A_\mu^z := \{x \in \mathbb{T}^z, M([\![\rho, x]\!]) = 0\}.$$

On s'intéressera aussi à l'ensemble R_μ des feuilles de $\mathbb{T}^z$ qui ne portent pas de mutations autres que celles de la branche origine sur leur lignée, et l'**arbre clonal réduit** qui est le sous arbre de $\mathbb{T}^z$ engendré par la racine et R_μ :

$$R_\mu := \{t \in \mathbb{R}_+, M_t(\mathbb{R}_+) = 0)\}.$$

Remarque 1. Cet ensemble R est étudié dans l'article de Philippe Marchal [8] pour un CPP avec $\nu = \frac{dx}{x^2}$ et des mutations à la naissance avec probabilité $1 - \alpha$. Dans ce cas les ensembles R_α ont la même loi que l'image d'un subordinateur α-stable. Dans le cas des mutations selon un processus de Poisson sur les branches, R n'est pas stable, mais il a une mesure de Lebesgue positive, ce qui le rend plus simple à caractériser (voir annexe A.1).

2.1.1 Nombre total de mutations

Maintenant que l'on a construit un arbre aléatoire muni de mutations aléatoires, on peut commencer à s'interroger sur la loi de cet objet. Commençons par une considération simple mais qui permet d'appréhender les difficultés qui peuvent apparaître avec cette construction. Comme on considère μ une mesure finie, le nombre de mutations sur une lignée fixée est une variable de Poisson de paramètre $\mu([0, z])$, donc est presque sûrement fini. Cependant, il se peut que dans un sous-arbre, il y ait un nombre infini de mutations avec probabilité 1. Par exemple, si μ est la mesure de Lebesgue et si ν l'intensité du CPP est telle que

$$\int_0 x\nu(\mathrm{d}x) = \infty,$$

on sait d'après les propriétés des mesures ponctuelles de Poisson (voir la propriété A.4 dans les annexes) que presque sûrement, la somme des longueurs des branches de tout sous-arbre complet (toutes les lignées issues d'un point) est infinie. Dans ce cas, le nombre de mutations de tout sous-arbre complet est infini. Alors de tout point de l'arbre, on peut choisir une mutation dans le sous-arbre des descendants, et itérer le procéder à l'infini pour trouver une lignée qui contient une infinité de mutations. Comme c'est vrai pour tout point de l'arbre, l'ensemble de ces lignées pathologiques est presque sûrement dense dans l'ensemble des lignées.

On se demande donc sous quelles conditions ce phénomène se produit. Conditionnellement à l'arbre de hauteur z, le nombre total de mutations suit une loi de Poisson de paramètre

$$\Lambda := \underline{\mu}(z) + \sum_{(t,y)\in\mathcal{N}, t<T(z)} \underline{\mu}(y),$$

où $T(z)$ est le premier temps tel qu'il y ait un atome de $\mathcal{N}$ au dessus de z. En effet, la branche origine est de hauteur z et les autres branches correspondent aux atomes de la mesure ponctuelle $\mathcal{N}$. Ce nombre de mutations est fini *p.s* sur l'événement $A := \{\Lambda < \infty\}$ et infini *p.s* sur le complémentaire de A. Or par les propriétés des mesures de Poisson, on distingue deux cas : soit A est de probabilité nulle, soit il est de probabilité 1.

Proposition 2.1. *Deux cas sont possibles,*

$$\int_0 \underline{\mu}(x)\nu(\mathrm{d}x) < \infty \quad \Longrightarrow \quad \textit{le nombre de mutations total est fini p.s.}$$
$$\int_0 \underline{\mu}(x)\nu(\mathrm{d}x) = \infty \quad \Longrightarrow \quad \textit{le nombre de mutations total est infini p.s.}$$

Démonstration. Conditionnellement à $T(z)$, l'ensemble $\mathcal{N}' := \{(t, y) \in \mathcal{N}, t < T(z)\}$ est un nuage Poissonnien sur $[0, T(z)] \times [0, z]$ d'intensité $\mathrm{d}t \otimes \nu$. Donc (d'après la propriété A.4) conditionnellement à $T(z)$, $\Lambda = \underline{\mu}(z) + \sum_{(t,y)\in\mathcal{N}'} \underline{\mu}(y)$ est fini *p.s* si et seulement si

$$T(z) \int_{[0,z]} \underline{\mu}(x) \wedge 1 \, \nu(\mathrm{d}x) < \infty,$$

et comme $T(z)$ est fini *p.s* et que $\underline{\mu}$ est croissante, cette condition est équivalente à la condition énoncée. □

Remarque 2. L'espérance du nombre total de mutations est

$$\mathbb{E}[\Lambda] = \underline{\mu}(z) + \frac{1}{\overline{\nu}(z)} \int_{[0,z]} \underline{\mu}(x)\nu(\mathrm{d}x).$$

2.2 Lien entre CPP et arbres de Yule

On considère le CPP de hauteur z_0 et d'intensité ν. Il est en général possible de le voir comme un arbre de Yule, modulo un changement de temps, à condition que les branchements soient forcément binaires.

Proposition 2.2. *Supposons ν sans atomes sur $\mathbb{R}_+$, avec $\overline{\nu}(z_0) > 0$. Si l'on effectue le changement de temps $t = \log \overline{\nu}(z)$, alors le CPP($\nu$, z_0) a la même loi qu'un arbre binaire où chaque individu donne naissance à taux 1 indépendamment des autres, démarré au temps $\log \overline{\nu}(z_0)$ avec un individu et arrêté au temps $\log \overline{\nu}(0)$ (éventuellement infini).*

Démonstration. Ceci est dû à la propriété de branchement de la mesure ponctuelle de Poisson $\mathcal{N}$ (voir l'annexe A.2.1). Cette propriété nous dit que l'arbre CPP(ν, z) peut être construit en tirant la hauteur Z du plus haut atome de $\mathcal{N}$, puis en recollant deux arbres indépendants CPP(ν, Z). Il suffit de constater que le taux de branchement est égal à 1 dans la nouvelle échelle de temps, or on connaît la loi de Z (premier point de branchement, voir l'annexe A.2.1) :

$$\mathbb{P}(\log \overline{\nu}(Z) > \log \overline{\nu}(z) + t) = \mathbb{P}(\overline{\nu}(Z) > \overline{\nu}(z)\mathrm{e}^t) = \int_{\overline{\nu}(z)\mathrm{e}^t} \frac{\overline{\nu}(z)}{u^2}\mathrm{d}u = \mathrm{e}^{-t}. \qquad \square$$

Mutations comme morts On peut en plus supposer que des mutations sont rajoutées selon un processus de Poisson d'intensité θ sur l'arbre original. Considérons-les comme des morts, ainsi l'on observe l'évolution de la population clonale en tant que processus de naissance et de mort. Dans le cas $\nu(\mathrm{d}x) = \frac{\mathrm{d}x}{x^2}$, on a alors un arbre de Yule auquel on rajoute un taux de mort $d(t) = \theta\mathrm{e}^{-t}$. En effet, le changement de temps est $t = -\log z$, et donc on exprime la probabilité qu'un individu ne meure pas entre les temps t et s :

$$\begin{aligned}\mathbb{P}(\text{pas de mort entre } t \text{ et } s) &= \mathbb{P}(\text{pas de mutation sur } [\mathrm{e}^{-s}, \mathrm{e}^{-t}]) \\ &= \mathrm{e}^{-\theta(\mathrm{e}^{-t}-\mathrm{e}^{-s})} \\ &= \exp\left(-\int_t^s \theta\mathrm{e}^{-u}\mathrm{d}u\right).\end{aligned}$$

Remarque 3. On connaît donc la loi de l'arbre clonal A_θ^z dans le cas $\nu = \frac{\mathrm{d}x}{x^2}$. C'est, modulo un changement de temps explicite, l'arbre simple associé à processus de naissance et de mort avec taux de branchement $r(t) = 1 + \mathrm{e}^{-t}$, avec N_t qui vaut 0 avec probabilité $\frac{\mathrm{e}^{-t}}{1+\mathrm{e}^{-t}}$ et 2 avec probabilité $\frac{1}{1+\mathrm{e}^{-t}}$. Autrement dit, le taux de naissance vaut 1 et le taux de mort vaut e^{-t}.

Remarque 4. On peut éviter de passer par un changement de temps. Alors on observe un processus de naissance et de mort avec le temps s'écoulant de z à 0. Par exemple, dans le cas $\nu = \frac{\mathrm{d}x}{x^2}$ et $\mu = \theta\mathrm{d}x$, quelques calculs permettent de dire que dans cette échelle, le taux de naissance est $b(x) = \frac{1}{x}$ et le taux de mort est simplement θ.

3 Construction couplée, processus Markovien

Une manière naturelle de poser des mutations le long des branches est de les lancer selon un processus de Poisson d'intensité fixe θ, c'est-à-dire que l'on considère $\mu = \theta \mathrm{d}x$. Il existe en outre une manière naturelle de coupler de tels processus de Poisson de façon à ce que les ensembles de mutations soient croissants en θ pour l'inclusion. En effet si M désigne une mesure ponctuelle de Poisson d'intensité la mesure de Lebesgue sur $\mathbb{R}_+^2$, on peut définir pour $\theta \geq 0$, $M^\theta := M([0, \theta] \cap \cdot)$. Alors M^θ est un processus de Poisson sur $\mathbb{R}_+$ d'intensité θ, et la suite $(M^\theta)_{\theta \geq 0}$ vue comme une suite d'ensembles, est croissante (voir figure 5).

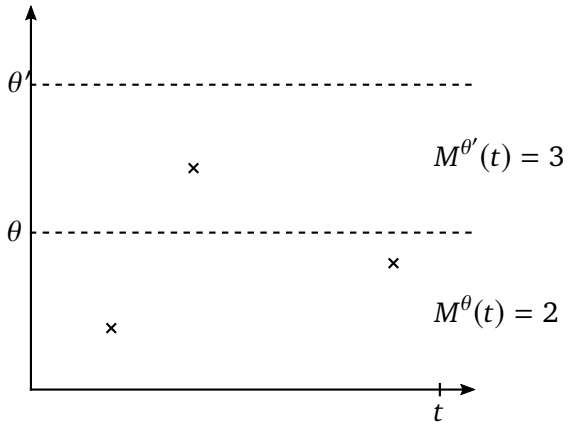

FIGURE 5 – Processus de Poisson couplés

C'est ce que l'on utilise pour construire des mutations couplées sur un arbre aléatoire. Pour chaque point $N = (t, x)$ de $\mathcal{N}$, on pose M_N une mesure ponctuelle de Poisson sur $\mathbb{R}_+ \times [0, x]$ d'intensité la mesure de Lebesgue. Pour θ fixé, on se ramène à la construction précédente avec $\mu = \theta \mathrm{d}x$ en considérant

$$M_N^\theta := M_N([0, \theta] \cap \cdot).$$

On a bien défini $(M^\theta)_{\theta \in \mathbb{R}_+}$ un couplage de mesures ponctuelles de Poisson sur l'arbre $\mathbb{T}^z$. Notons A_θ^z l'arbre clonal de hauteur z au niveau de mutation θ, c'est-à-dire le sous arbre réel

$$A_\theta^z := \{x \in \mathbb{T}^z, M^\theta([|\rho, x|]) = 0\}.$$

Le processus $(A_\theta^z)_{\theta \in \mathbb{R}_+}$ est un processus décroissant (au sens de l'inclusion) d'arbres réels, que l'on cherche à étudier.

3.1 Bourgeons, greffes

On a donc un arbre $\mathbb{T}^z$ de loi CPP(ν, z), muni d'un couplage de mutations $(M^\theta)_{\theta \in \mathbb{R}_+}$, qui permettent de définir le processus décroissant des arbres clonaux $(A_\theta^z)_{\theta \in \mathbb{R}_+}$ à valeurs dans les arbres réels simples. Dans le sens θ croissant, c'est clairement un processus de Markov tel que la loi de $A_{\theta+\theta'}^z$ sachant A_θ^z est la loi de l'arbre clonal obtenu en rajoutant des mutations au taux θ' le long des branches de A_θ^z. Dans le sens θ décroissant, le processus de l'arbre est croissant. On va chercher à décrire la façon dont cet arbre croît, en introduisant la notion de greffes d'arbres simples.

Définition 5. Pour un arbre réel simple $A = (\mathcal{T}, (\alpha(u), \zeta(u), \omega(u))_{u\in\mathcal{T}})$, on définit les **bourgeons** de A comme l'ensemble $\mathcal{B}(A)$ des feuilles de l'arbre discret $\mathcal{T}$ qui n'ont pas une vie infinie :

$$\mathcal{B}(A) := \{b \in \mathcal{T},\ b1 \notin \mathcal{T},\ \omega(b) < \infty\}.$$

Pour deux arbres simples $A_i = (\mathcal{T}_i, (\alpha_i(u), \zeta_i(u), \omega_i(u))_{u\in\mathcal{T}_i})$ avec $i \in \{1, 2\}$, et pour $b \in \mathcal{B}(A_1)$, on peut définir la **greffe** de A_2 sur A_1 au niveau du bourgeon b, notée $A_1 \oplus_b A_2$, en recollant la racine de A_2 sur la feuille qui correspond à la fin de la vie de b dans A_1. Formellement, on définit :

$$\mathcal{T} := \mathcal{T}_1 \cup b\mathcal{T}_2,$$

$$\alpha(b) := \alpha_1(b), \quad \omega(b) := \omega_1(b) + \zeta_2(\emptyset),$$

$$\forall u \in \mathcal{T}_1 \setminus \{b\}, \quad \alpha(u) := \alpha_1(u), \quad \omega(u) := \omega_1(u),$$

$$\forall u \in \mathcal{T}_2 \setminus \{\emptyset\}, \quad \begin{cases} \alpha(bu) := \omega(b) + (\alpha_2(u) - \omega_2(\emptyset)), \\ \omega(bu) := \alpha(bu) + \zeta_2(u), \end{cases}$$

$$A_1 \oplus_b A_2 := (\mathcal{T}, (\alpha(u), \zeta(u), \omega(u))_{u\in\mathcal{T}}).$$

On a alors clairement $\mathcal{B}(A_1 \oplus_b A_2) := \mathcal{B}(A_1) \setminus \{b\} \cup b\mathcal{B}(A_2)$.

Remarque 5. En fait, on peut toujours greffer la racine d'un arbre réel sur un autre pour former un nouvel arbre réel, mais cette greffe sur les bourgeons a l'avantage de conserver la structure des arbres simples (voir figure 6).

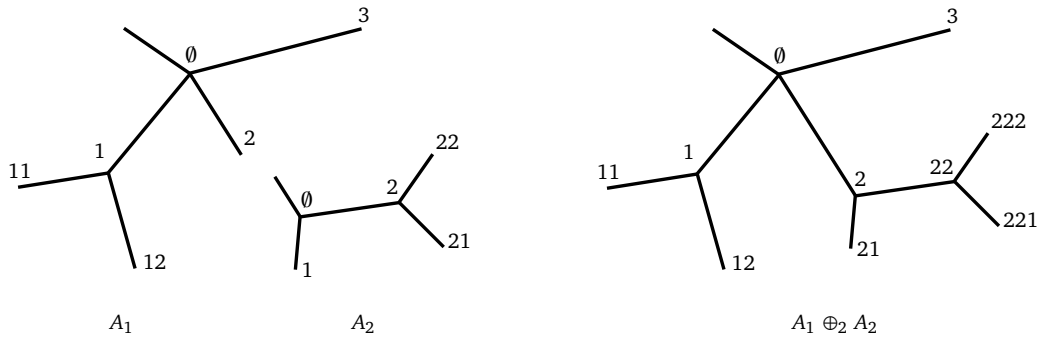

FIGURE 6 – Greffe d'arbres simples

On aura besoin de faire des greffes un peu plus générales. Définissons donc les greffes récursives.

Définition 6. Soit A un arbre simple, B un sous-ensemble de $\mathcal{U}$, et $(A_b)_{b\in B}$ des arbres simples. Pour $b \in B$ on pose $\{b_1, \ldots, b_{n(b)}\} = \{b' \in B,\ b' \prec b\}$, indexés tels que $b_1 \prec b_2 \prec \ldots \prec b_{n(b)} = b$.
Supposons que pour tout $b \in B$, pour tout $1 \le i \le n(b)$, b_i est bien un bourgeon de $A \oplus_{b_1} A_{b_1} \oplus_{b_2} \cdots \oplus_{b_{i-1}} A_{b_{i-1}}$.
Alors on peut bien définir la greffe récursive des (A_b) sur A en posant

$$A \oplus_B (A_b)_{b\in B} := \bigcup_{b\in B} A \oplus_{b_1} A_{b_1} \oplus_{b_2} \cdots \oplus_b A_b,$$

où l'union est à comprendre comme une limite d'arbres simples.

On admet sans trop expliciter l'arbre simple final que de telles greffes sont bien définies.

Remarque 6. Ici, nos arbres poussent de z (la hauteur de la racine) jusqu'à 0 (l'axe des abscisses). Quitte à inverser l'axe des ordonnées, on s'autorise à les considérer comme des arbres simples, que l'on peut greffer les uns aux autres.

Pour étudier l'arbre qui pousse, on va donc retourner le temps en posant $\eta = -\log\theta$, et en étudiant le processus croissant $(X_\eta^z := A^z_{\mathrm{e}^{-\eta}})_{\eta\in\mathbb{R}}$. Notons $\mathbb{Q}_\eta^z$ la loi de X_η^z à support dans les arbres simples binaires. Ce processus admet en fait une description plutôt agréable.

Proposition 3.1. *Le processus $(X_\eta^z)_{\eta\in\mathbb{R}}$ est un processus de Markov inhomogène à valeurs dans les arbres simples. Les bourgeons de ces arbres sont les feuilles de hauteur strictement positives (qui correspondent aux mutations au niveau $\theta = \mathrm{e}^{-\eta}$). Indépendamment des autres, chaque bourgeon b, de hauteur $\omega(b)$, "pousse" au bout d'un temps exponentiel de paramètre* 1. *Le sens de "pousser" au temps η est ici : on greffe un arbre sur le bourgeon b, qui suit la loi $\mathbb{Q}_\eta^{\omega(b)}$ (voir figure 7).*

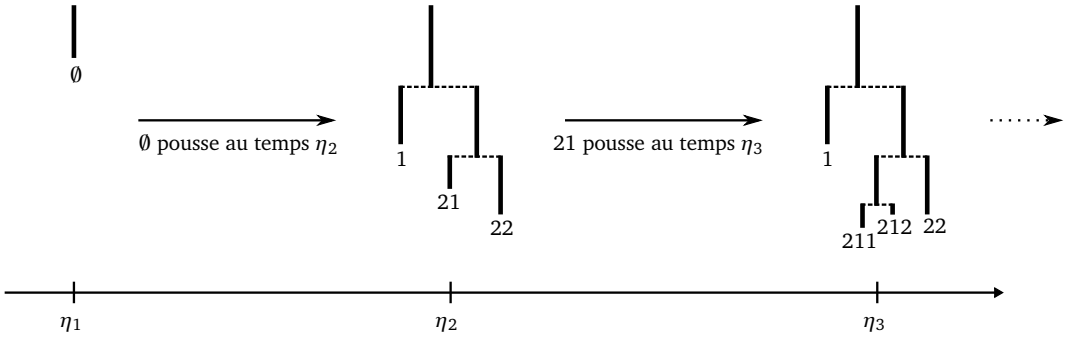

FIGURE 7 – Évolution Markovienne d'un arbre croissant. Dans cet exemple, le temps $\eta_2 - \eta_1$ est un temps exponentiel de paramètre 1 et $\eta_3 - \eta_2$ est un temps exponentiel de paramètre 3.

3.2 Branchements d'arbres simples

Montrons qu'un tel processus qui évolue par greffes consécutives est bien défini. On cherche donc à construire un processus $(X'^z_\eta)_{\eta\in\mathbb{R}}$ qui évolue comme dans la proposition 3.1. Construisons un arbre discret $\mathcal{T}$ qui va hiérarchiser l'ensemble des bourgeons du processus au cours du temps. C'est-à-dire que l'on va définir des variables aléatoires $(b_u, Z_u, \xi_u, \eta_u, Y_u)_{u\in\mathcal{T}}$ indexées par $\mathcal{T}$. Ainsi chaque $u \in \mathcal{T}$ représentera un bourgeon b_u du processus final, qui vit à la hauteur $Z_u > 0$ pendant un temps $\xi_u > 0$ et devient à l'instant $\eta_u \in \mathbb{R}$ un arbre simple Y_u.

Fixons tout d'abord $(\xi_u)_{u\in\mathcal{U}}$ des variables *i.i.d* exponentielles de paramètre 1. Soit $\eta_0 \in \mathbb{R}$ et $z > 0$ fixés.

— **Initialisation** : on pose $Z_\emptyset := z$, $\eta_\emptyset := \eta_0$, et on définit $\mathcal{T}^1 = \{\emptyset\}$.

— **Exploration** : à l'étape n, pour chaque feuille u de $\mathcal{T}^n$, on définit l'arbre Y_u selon la loi $\mathbb{Q}_{\eta_u}^{Z_u}$, indépendamment de tout le reste. Cet arbre un nombre N_u éventuellement infini de bourgeons $\{b_1, b_2, \ldots\}$, à des hauteurs $\omega_{Y_u}(b_1)$, $\omega_{Y_u}(b_2)$, ...

On pose pour $i \leq N$,

$$b_{ui} = b_u b_i, \quad Z_{ui} := \omega_{Y_u}(b_i), \quad \eta_{ui} := \eta_u + \xi_{ui}.$$

Enfin on pose

$$\mathcal{T}^{n+1} := \mathcal{T}^n \cup \bigcup_{u \text{ feuille de } \mathcal{T}^n} \{ui, 1 \leq i \leq N_u\}.$$

— On recommence la phase d'exploration, génération par génération, et on décrit bien tout un sous-arbre $\mathcal{T} := \cup_n \mathcal{T}^n$ de $\mathcal{U}$.
— Pour tous les sommets non explorés de $\mathcal{U} \setminus \mathcal{T}$, on peut poser $\eta_u = \infty$ pour des raisons pratiques.

3.2.1 Construction de la croissance de l'arbre

On cherche simplement à formaliser la description de la proposition. Le processus $(X'^z_\eta)_{\eta \geq \eta_0}$ sera donc le recollement de tous les arbres Y_u, pour tous les $u \in \mathcal{T}$ tels que $\eta_u \leq \eta$. C'est-à-dire, que pour $\eta \geq \eta_0$, on pose

$$X'^z_\eta = Y_\emptyset \oplus_{B_\eta} \{Y_u,\ u \in \mathcal{T},\ \eta_u \leq \eta\},$$

avec $B_\eta = \{b_u,\ u \in \mathcal{T},\ \eta_u \leq \eta\}$ l'ensemble de bourgeons qui convient.

Finalement on peut montrer ce qui nous intéresse : le processus $(X'^z_\eta)_{\eta \geq \eta_0}$ ainsi défini a la même loi que le processus de l'arbre clonal $(X^z_\eta)_{\eta \geq \eta_0}$, ce qui prouve la proposition 3.1.

Idée de preuve. Par définition, $X'^z_{\eta_0}$ et $X^z_{\eta_0}$ suivent la même loi $\mathbb{Q}^z_{\eta_0}$. Montrons que cela reste vrai pour tout le processus.

Soit $\eta_0 \leq \eta_1 < \eta_2$. On peut reformuler la construction jointe de $(X^z_{\eta_1}, X^z_{\eta_2})$. On a un arbre de naissance et de mort avec un temps allant de z à 0, chaque branchement arrivant à taux $\frac{1}{x}$ avec le long des branches un processus ponctuel de Poisson qui donne l'emplacement des mutations, et leur niveau $\theta \geq 0$. Alors l'arbre $X^z_{\eta_1}$ est l'arbre arrêté aux premières mutations de niveau $\theta \leq \mathrm{e}^{-\eta_1}$, et l'arbre $X^z_{\eta_2}$ est le même arbre arrêté aux premières mutations de niveau $\theta \leq \mathrm{e}^{-\eta_2}$. Vu la construction des mutations comme processus ponctuel de Poisson d'intensité la mesure de Lebesgue, chaque mort pour $X^z_{\eta_1}$ a une probabilité $\mathrm{e}^{-\eta_2}/\mathrm{e}^{-\eta_1} = \mathrm{e}^{-(\eta_2-\eta_1)}$ d'être également une mort pour $X^z_{\eta_2}$. De plus, depuis les morts (disons à la hauteur z') de $X^z_{\eta_1}$ qui n'en sont pas pour $X^z_{\eta_2}$, par la propriété de branchement le sous-arbre émergent a la loi $\mathbb{Q}^{z'}_{\eta_2}$. Pour résumer schématiquement, on a (de manière indépendante pour chaque b)

$$X^z_{\eta_2} = X^z_{\eta_1} \bigoplus_{b \in \mathcal{B}(X^z_{\eta_1})} \begin{cases} Y_b \sim \mathbb{Q}^{\omega(b)}_{\eta_2} & \text{avec probabilité } 1 - \mathrm{e}^{-(\eta_2-\eta_1)}, \\ \emptyset & \text{avec probabilité } \mathrm{e}^{-(\eta_2-\eta_1)}. \end{cases}$$

On a donc bien, indépendamment du reste de l'arbre, sachant que b est un bourgeon de $X^z_{\eta_1}$, le premier temps $\eta' \geq \eta_1$ tel que l'on greffe un arbre sur b est tel que $\eta' - \eta_1$ est une variable exponentielle de paramètre 1. Aussi, on sait qu'à tout moment $\eta_2 \geq \eta'$, l'arbre greffé a loi $\mathbb{Q}^{\omega(b)}_{\eta_2}$. En particulier, l'arbre greffé à l'instant η' suit la loi $\mathbb{Q}^{\omega(b)}_{\eta'}$. Finalement, on se convainc que le processus a la même loi que le processus (X'^z_η), construit explicitement avec des greffes d'arbres simples. □

3.3 Générateur infinitésimal

À partir de $\eta \in \mathbb{R}_+$ et A un arbre simple fixés, si $X_\eta^z = A$, chaque bourgeon b de A peut éclore indépendamment des autres au taux 1 après le temps η. Pour n'importe quel temps η' qui suit l'éclosion, l'arbre que l'on a greffé au bourgeon b suit la loi $\mathbb{Q}_{\eta'}^{\omega(b)}$. Pour $\epsilon > 0$ fixé, à tout moment $\eta > 0$, l'arbre X_η^z a un nombre fini de bourgeons au-dessus du niveau ϵ, c'est-à-dire des feuilles b telles que $\omega(b) > \epsilon$. Si ϕ est une fonction des arbres réels (simples) qui ne dépend que du sous-arbre à une distance inférieure à $z - \epsilon$ de la racine, on voit bien que le générateur infinitésimal $\mathcal{L}_\eta$ du processus $(X_\eta^z)_{\eta \geq \eta_0}$ vérifie,

$$\mathcal{L}_\eta \phi(A) = \sum_{b \in \mathcal{B}(A)} \left(\mathbb{Q}_\eta^{\omega(b)}[\phi(A \oplus_b Y)] - \phi(A) \right).$$

Si l'on oublie la structure de l'arbre et que l'on considère chaque bourgeon, repéré par sa hauteur, indépendamment des autres, on obtient une marche branchante assez simple. Soit $\chi_\eta := \sum_{b \in \mathcal{B}(X_\eta^z)} \delta_{\omega(b)}$. Alors $(\chi_\eta)_{\eta \geq \eta_0}$ est un processus de branchement où les individus restent immobiles (à leur hauteur z) pendant leur temps de vie, puis branchent lors de leur mort au temps η selon la loi des hauteurs des bourgeons de l'arbre Y sous la mesure $\mathbb{Q}_\eta^z$. On peut décrire quelques espérances de fonctionnelles de cette mesure : soit $f : \mathbb{R}_+ \mapsto \mathbb{R}_+$, nulle au voisinage de 0. Alors on a, avec $\theta = \mathrm{e}^{-\eta}$,

$$\mathbb{Q}_\eta^z(\chi(f)) = \int_0^z \frac{z}{s} \theta \mathrm{e}^{-\theta(z-s)} f(s) \mathrm{d}s.$$

Démonstration. On pose $F(z) = \mathbb{Q}_\eta^z(\chi(f))$. Alors par la propriété de branchement de l'arbre (dans l'échelle du temps x qui va de z à 0, avec taux de naissance en $1/x$, taux de mort en θ),

$$F(z) = \mathrm{e}^{-\theta z} \phi(0, z) + \int_0^z \theta \mathrm{e}^{-\theta(z-s)} (f(s) + \phi(s, z)) \mathrm{d}s,$$

avec

$$\phi(s, z) := \int_s^z \frac{1}{u} F(u) \mathrm{d}u.$$

Donc (on dérive), F vérifie une équation différentielle :

$$\begin{cases} F'(z) = -\theta F(z) + \frac{1}{z} F(z) + \theta f(z), \\ F(0) = 0. \end{cases}$$

Autrement dit, F s'écrit

$$F(z) = \mathrm{e}^{-\theta z + \log z} \int_0^z \theta f(s) \frac{\mathrm{e}^{\theta s}}{s} \mathrm{d}s. \qquad \square$$

4 Population clonale, partition allélique

4.1 Ensemble régénératif des lignées clonales, CPP clonal

On rappelle que l'on construit un CPP(ν, z), c'est-à-dire un arbre ultramétrique aléatoire de hauteur z, à partir d'un processus ponctuel de Poisson $\mathcal{N}$ d'intensité

$\mathrm{d}t \otimes \nu$, avec ν sans atomes. On a également jeté des mutations le long des branches de l'arbre selon des processus de Poisson indépendants d'intensité μ. On se propose d'étudier l'ensemble des lignées (feuilles de l'arbre) qui ne portent pas de mutation. Posons donc :

$$R = \{t \geq 0,\ \ M_t(\mathbb{R}_+) = 0\}.$$

On étudie aussi l'arbre clonal réduit, c'est-à-dire le sous-arbre de l'arbre original engendré par la racine et les lignées clonales (les points de R). On supposera dans cette partie que **la branche ancestrale (l'axe des ordonnées) ne porte pas de mutations**, ainsi on a toujours $0 \in R$. Cet ensemble est régénératif (voir annexe A.1), on peut essayer de le caractériser par son exposant de Laplace.

Proposition 4.1. *Sous les hypothèses (H) et avec les notations précédentes, l'ensemble aléatoire des lignées clonales de l'origine R est régénératif. On peut le caractériser par l'exposant de Laplace du subordinateur dont il est l'image :*

$$\frac{1}{\phi(\lambda)} = \int_{]0,\infty[} \frac{\mathrm{e}^{-\underline{\mu}(x)}}{\lambda + \overline{\nu}(x)} \mu(\mathrm{d}x),$$

où $\overline{\nu}(x) = \nu([x, \infty])$ et $\underline{\mu}(x) = \mu([0, x])$.

Aussi, l'arbre clonal réduit (le sous-arbre engendré par les feuilles dans l'ensemble R) a la loi d'un CPP d'intensité ν^μ une mesure sur $\mathbb{R}_+ \cup \{\infty\}$. En posant $W(x) := (\overline{\nu}(x))^{-1}$ et $W^\mu(x) := (\overline{\nu^\mu}(x))^{-1}$, on a la caractérisation suivante. Pour tout $x > 0$ tel que $W(x) < \infty$,

$$W^\mu(x) = W(0) + \int_0^x \mathrm{e}^{-\underline{\mu}(z)} \mathrm{d}W(z).$$

Remarque 7. La dernière formule de la proposition est une extension de la proposition 3.1 de l'article d'Amaury Lambert [6], où le cas ν mesure finie et $\mu = \theta \mathrm{d}x$ est traité. On permet ici que ν soit de masse infinie, et μ peut prendre une forme générale (à condition que $\underline{\mu}(x) < \infty$ pour un $x > 0$).

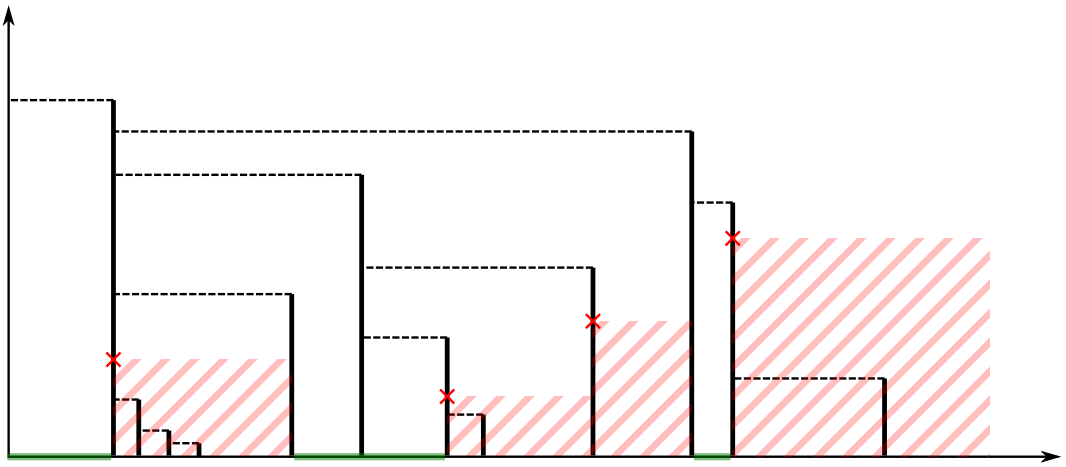

FIGURE 8 – CPP muni de mutations, ensemble régénératif R représenté en vert

Ensemble régénératif On démontre d'abord la première partie de la proposition sur l'ensemble régénératif R.

Démonstration du début de la proposition 4.1. Soit $(\mathcal{F}_t)_{t\geq 0}$ la filtration naturelle du CPP marqué, définie par

$$\mathcal{F}_t = \sigma\left(\mathcal{N} \cap ([0,t]\times\mathbb{R}_+),\, M_{(s,x)},\, s\leq t,\, x\geq 0\right).$$

Pour montrer que R est bien $(\mathcal{F}_t)$-progressivement mesurable, on montre que pour $t>0$ fixé, l'application

$$\begin{cases} [0,t]\times\Omega & \longrightarrow \mathbb{R}_+ \\ (s,\omega) & \longmapsto \mathbf{1}_{s\in R(\omega)} \end{cases}$$

est $\mathcal{F}_t$-mesurable. Soit (U_i, X_i) une suite des coordonnées des mutations d'abscisses U_i inférieures à t (par exemple classée par ordre des ordonnées X_i décroissantes, voir figure 9). Ces variables aléatoires peuvent être choisies de façon $\mathcal{F}_t$-mesurables, ainsi que les variables

$$T_i := t \wedge \inf\{s\geq U_i,\, (s,x)\in\mathcal{N},\, x\geq X_i\}.$$

Alors comme on a

$$R\cap[0,t] = \bigcap_i ([0,t]\setminus[U_i, T_i[),$$

R est bien $(\mathcal{F}_t)$-progressivement mesurable.

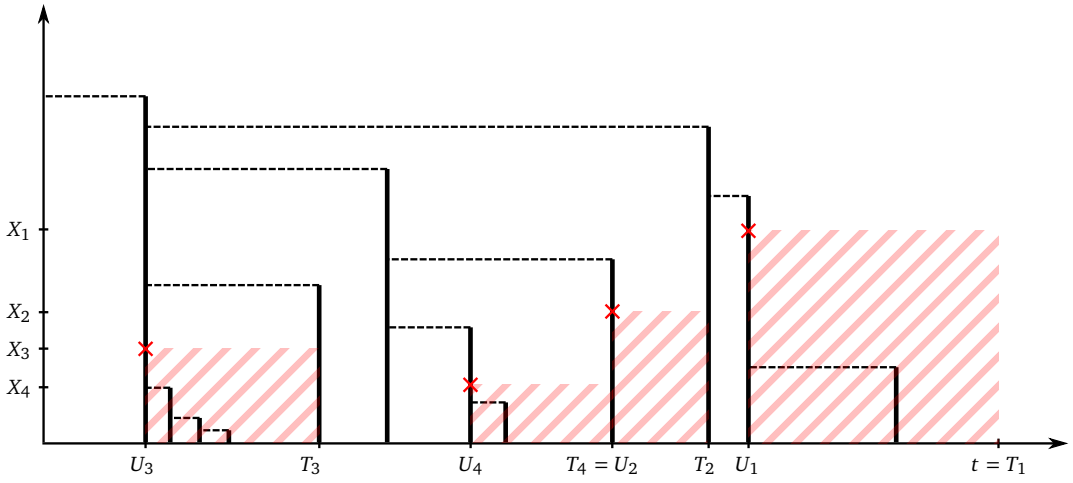

FIGURE 9 – Mutations localisées par les variables (U_i, X_i, T_i)

On va maintenant montrer la propriété de régénération de R. Soit

$$H(s,t) := \max\{x\geq 0,\, (u,x)\in\mathcal{N},\, s<u\leq t\},$$

la hauteur maximale des atomes de $\mathcal{N}$ entre s et t. On notera $H(t) := H(0,t)$ pour simplifier. Remarquons que

$$R = \{t\geq 0,\, M_t([0,H(t)]) = 0\}.$$

Soit un $(\mathcal{F}_t)$-temps d'arrêt T tel que presque sûrement $T < \infty$ et $T \in R$ est non isolé à droite. Par les propriétés des processus de Poisson, et le fait que les variables $(M_{(s,x)})_{s\geq 0, x\geq 0}$ sont *i.i.d.*, on sait que l'arbre strictement à droite de T est indépendant de $\mathcal{F}_T$ et a la même loi que l'arbre initial. Or comme $T \in R$ presque sûrement, a

$$R \cap [T, \infty[= \{t \geq T, M_t([0, H(T, t)]) = 0\}.$$

Donc il est clair que $R \cap [T, \infty[- T$ a la même loi que R et est indépendant de $\mathcal{F}_T$.

R est donc régénératif, on peut calculer son exposant de Laplace. Ici, R a une mesure de Lebesgue non nulle, en particulier on a pour $t \in \mathbb{R}_+$,

$$\begin{aligned}
\mathbb{P}(t \in R) &= \mathbb{E}\left[\mathrm{e}^{-\underline{\mu}(H_t)}\right] \\
&= \int_{[0,\infty]} \mathbb{P}(H_t \in \mathrm{d}x)\mathrm{e}^{-\underline{\mu}(x)} \\
&= \int_{]0,\infty[} \mathbb{P}(H_t \leq x)\mathrm{e}^{-\underline{\mu}(x)}\mu(\mathrm{d}x) \\
&= \int_{]0,\infty[} \mathrm{e}^{-t\overline{\nu}(x)-\underline{\mu}(x)}\mu(\mathrm{d}x).
\end{aligned}$$

Le passage de la deuxième à la troisième ligne se fait par intégration par partie, grâce à l'hypothèse que $\underline{\mu}$ est continue et que μ est de masse infinie. C'est donc la densité par rapport à la mesure de Lebesgue de sa mesure de renouvellement. Cela suffit à caractériser notre ensemble régénératif, et en calculant la transformée de Laplace de cette mesure, on trouve bien l'expression de la proposition. □

Remarque 8. Dans le cas particulier $\nu = \frac{\mathrm{d}x}{x^2}$ et $\mu = \theta \mathrm{d}t$, on a

$$\frac{1}{\phi_\theta(\lambda)} = \int_0^\infty \frac{\theta \mathrm{e}^{-\theta x}}{\lambda + 1/x}\mathrm{d}x.$$

En particulier comme pour tout θ, $c > 0$, on a

$$\phi_\theta(c\lambda) = c\phi_{\theta/c}(\lambda),$$

on en déduit l'égalité en loi $cR_\theta \overset{(d)}{=} R_{\theta/c}$. Il y a donc une "quasi-stabilité" de la loi des ensembles régénératifs $(R_\theta)_\theta$ (ce n'est pas la manière la plus directe de montrer cette égalité en loi). Cependant R_θ n'est pas un ensemble régénératif stable comme le sont les R_α dans l'article de Marchal [8].

Arbre clonal Pour démontrer que l'arbre clonal réduit est un CPP, on met en évidence un nuage de points qui l'engendre. En effet, étant donné σ le subordinateur d'image R que l'on a construit, on pose

$$\mathcal{N}' := \{(t, x), t \in \mathbb{R}_+, x = H(\sigma_{t-}, \sigma_t) > 0\},$$

où $H(s, t) := \max\{x, (u, x) \in \mathcal{N}, s \leq u \leq t\}$. Cet ensemble de points définit bien le arbre clonal réduit, car $H(\sigma_{t-}, \sigma_t)$ est la distance dans l'arbre entre les feuilles consécutives $sigma_{t-}$ et σ_t de R. Pour achever la démonstration de la proposition, il suffit de montrer que conditionnellement au temps de mort ζ de σ, $\mathcal{N}'$ est un nuage Poissonnien sur $[0, \zeta[\times\mathbb{R}_+$ d'intensité $\mathrm{d}t \otimes \nu^\mu$.

Fin de la démonstration de la proposition 4.1. Ceci est dû à la propriété de régénération du processus. Pour $t \geq 0$ fixé, σ_t est un $(\mathcal{F}_t)$-temps d'arrêt, tel que sur $\{\sigma_t < \infty\} = \{\zeta > t\}$, σ_t est presque sûrement dans R. Ceci implique que conditionnellement à $\{\sigma_t < \infty\}$, le CPP marqué strictement à droite de σ_t est égal en loi au CPP marqué original, et est indépendant de $\mathcal{F}_{\sigma_t}$. En particulier :

$$\left(\{(s,x) \in \mathbb{R}_+^2, (\sigma_t + s, x) \in \mathcal{N}\}, R \cap [\sigma_t, \infty[- \sigma_t\right) \overset{(d)}{=} (\mathcal{N}, R).$$

Donc on remarque que $\mathcal{N}' \cap ([t,\infty[\times \mathbb{R}_+) - (t,0)$ à la même loi que $\mathcal{N}'$ et est indépendant de $\mathcal{F}_{\sigma_t}$. En fixant $\epsilon > 0$, on pose $(T_i, X_i)_i$ la suite des atomes de $\mathcal{N}'$ tels que $X_i > \epsilon$, rangés par ordre de T_i croissant. Alors T_i est un $(\mathcal{F}_{\sigma_t})$-temps d'arrêt et vu la remarque que l'on vient de faire, la suite $(T_i - T_{i-1}, X_i)_i$ est *i.i.d.* Il suffit de constater que la variable T_1 suit une loi exponentielle, ce qui découle aussi de la remarque, pour prouver que $\mathcal{N}'$ a une intensité de la forme $\mathrm{d}t \otimes \nu^\mu$.

Il reste à caractériser ν^μ en calculant $W^\mu(x)$. On note que les calculs qui suivent sont corrects grâce à l'hypothèse que ν est sans atome, donc que W est continue. Pour simplifier, posons $H_t := H(0,t) = \max\{x,\ (u,x) \in \mathcal{N},\ 0 \leq u \leq t\}$.

$$\begin{aligned} W^\mu(x) &= \int_0^\infty \mathrm{e}^{-t\overline{\nu^\mu}(x)}\mathrm{d}t \\ &= \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\{H_{\sigma_t} \leq x\}}\mathrm{d}t\right] \\ &= \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\{H_u \leq x\}}\mathbf{1}_{\{u \in R\}}\mathrm{d}u\right]. \end{aligned}$$

Or si l'on pose $F(y) := \mathbb{P}(H_u \leq y) = \mathrm{e}^{-u\overline{\nu}(y)}$, on a

$$\begin{aligned} \mathbb{P}(H_u \leq x, u \in R) &= \mathbb{P}(H_u = 0) + \int_0^x \mathbb{P}(H_u \in \mathrm{d}y)\mathrm{e}^{-\underline{\mu}(y)} \\ &= F(0) + \int_0^x \mathrm{e}^{-\underline{\mu}(y)}\mathrm{d}F(y). \end{aligned}$$

Mais $\mathrm{d}F(y) = u\mathrm{e}^{-u\overline{\nu}(y)}\nu(\mathrm{d}y)$, d'où

$$\begin{aligned} W^\mu(x) &= \int_0^\infty \mathrm{e}^{-u\overline{\nu}(0)}\mathrm{d}u + \int_0^x \left(\int_0^\infty u\mathrm{e}^{-u\overline{\nu}(y)}\mathrm{d}u\right)\mathrm{e}^{-\underline{\mu}(y)}\nu(\mathrm{d}y) \\ &= \frac{1}{\overline{\nu}(0)} + \int_0^x \frac{1}{\overline{\nu}(y)^2}\mathrm{e}^{-\underline{\mu}(y)}\nu(\mathrm{d}y) \\ &= W(0) + \int_0^x \mathrm{e}^{-\underline{\mu}(y)}\mathrm{d}W(y). \end{aligned}$$

□

Remarque 9. L'égalité de la première équation de la preuve devient, quand x tend vers l'infini :

$$W^\mu(\infty) = \mathbb{E}[\mathrm{Leb}(R)].$$

Remarque 10. Dans notre cas $\nu = \frac{\mathrm{d}x}{x^2}$ et $\mu = \theta \mathrm{d}x$, on a

$$W^\theta(x) = \frac{1 - \mathrm{e}^{-\theta x}}{\theta}.$$

L'arbre clonal est donc engendré par les hauteurs des excursions d'un Brownien avec drift $\theta/2$, ce qui est cohérent avec les résultats d'Abraham et Serlet sur le *Poisson Snake* [1].

4.2 Taille de la population clonale

Comme l'ensemble des feuilles de l'arbre est identifié à l'axe des abscisses, on dispose d'une mesure naturelle sur cet ensemble, la mesure de Lebesgue. On voudrait calculer la loi de $\mathrm{Leb}(R)$, la taille de la population clonale. On déduit facilement des constructions précédentes le résultat suivant.

Corollaire 4.2. *Sous les hypothèses* (H) *et avec les notations de la proposition 4.1, la taille de la population clonale $\mathrm{Leb}(R)$ est une variable exponentielle d'espérance $W^\mu(\infty)$. Dans un CPP(ν, z) où l'on ignore les mutations sur la branche origine, la taille de la population clonale est une variable exponentielle d'espérance $W^\mu(z)$.*

Démonstration. Selon notre construction du subordinateur σ d'image R, on a

$$\mathrm{Leb}(R) = L_\infty = \inf\{t > 0, \sigma_t = \infty\}.$$

On sait que c'est une variable aléatoire de loi exponentielle, de paramètre $\phi(0)$, où ϕ est l'exposant de Laplace de σ. Dans le cadre général, on a donc

$$\phi(0)^{-1} = \mathbb{E}\left[\mathrm{Leb}(R)\right] = W^\mu(\infty).$$

Si l'on fixe une hauteur $z > 0$, on s'intéresse à la loi de $\mathrm{Leb}(R \cap [0, T(z)])$, où $T(z)$ est le premier temps tel qu'il y ait une branche plus grande que z. C'est la taille de la population clonale dans le CPP de hauteur z. Par les propriétés des mesures ponctuelles de Poisson, cela revient à étudier la population clonale totale d'un CPP d'intensité ν' et de même taux μ de mutations, avec

$$\nu' = \nu(\,\cdot\, \cap [0, z]) + \overline{\nu}(z)\delta_\infty.$$

Alors si $W'(x) := \overline{\nu}'(x)^{-1}$, on a

$$W'(x) = W(x \wedge z),$$

et alors $(W')^\mu(\infty) = W^\mu(z)$. Donc $\mathrm{Leb}(R \cap [0, T(z)])$ suit une loi exponentielle d'espérance $W^\mu(z)$. □

4.3 Probabilité qu'un clone existe

Dans cette partie, on considère le CPP d'intensité ν et de hauteur z, pourvu de mutations le long des branches selon la mesure μ, cette fois également sur la branche ancestrale. Alors avec probabilité $\mathrm{e}^{-\mu(z)}$, il n'y a pas de mutations sur la branche de l'origine et on se retrouve dans le cas particulier précédent. Sinon il peut ne pas y avoir d'arbre clonal, ou il est possible que la population clonale existe mais à une distance strictement positive de l'origine. On voudrait calculer la probabilité qu'il existe un clone, ce qui se déduit assez bien des calculs précédents.

Proposition 4.3. *Dans un CPP(ν, z) muni de mutations selon une mesure μ, sous les hypothèses* (H) *et avec les notations de la proposition 4.1, il existe un clone, c'est-à-dire une lignée qui ne porte pas de mutations, avec probabilité*

$$\frac{W(z)\mathrm{e}^{-\underline{\mu}(z)}}{W^{\mu}(z)}.$$

Démonstration. Pour prouver cela, on utilise la propriété de régénération de notre construction : soit X le premier clone sur l'axe des abscisses. X est un $(\mathcal{F}_t)$-temps d'arrêt, et conditionnellement à $\{X < \infty\}$, la loi de l'arbre à droite de X est la même que la loi de l'arbre original conditionné à ne pas avoir de mutations sur la branche des origines. On note $C = \{X < \infty\}$ l'événement *il existe un clone* et M l'événement *il n'y a pas de mutation sur la branche des origines*. On a alors

$$\begin{aligned}\mathbb{E}\left[\mathrm{Leb}(R)\right] &= \mathbb{P}(C)\mathbb{E}\left[\mathrm{Leb}(R) \mid C\right] \\ &= \mathbb{P}(C)\mathbb{E}\left[\mathrm{Leb}(R \cap [X, \infty[- X) \mid X < \infty\right] \\ &= \mathbb{P}(C)\mathbb{E}\left[\mathrm{Leb}(R) \mid M\right]\end{aligned}$$

Il suffit donc de calculer les deux quantités $\mathbb{E}\left[\mathrm{Leb}(R)\right]$ et $\mathbb{E}\left[\mathrm{Leb}(R) \mid M\right]$ pour connaître la probabilité de l'événement C. On a déjà calculé la deuxième quantité en cherchant à exprimer W^{μ} dans la section précédente :

$$\mathbb{E}\left[\mathrm{Leb}(R) \mid M\right] = \int_0^{\infty} \mathbb{P}(t \in R,\, H_t < z \mid M)\,\mathrm{d}t = W^{\mu}(z).$$

Il reste à calculer l'autre terme :

$$\begin{aligned}\mathbb{E}\left[\mathrm{Leb}(R)\right] &= \mathbb{E}\int_0^{T(z)} \mathbf{1}_{\{t \in R\}}\,\mathrm{d}t \\ &= \int_0^{\infty} \mathbb{P}(t \in R,\, t < T(z))\,\mathrm{d}t \\ &= \int_0^{\infty} \mathrm{e}^{-t\overline{\nu}(z)}\mathrm{e}^{-\underline{\mu}(z)}\,\mathrm{d}t \\ &= \frac{\mathrm{e}^{-\underline{\mu}(z)}}{\overline{\nu}(z)} = W(z)\mathrm{e}^{-\underline{\mu}(z)}.\end{aligned}$$

D'où la probabilité qu'il existe un clone de l'origine dans la population au temps présent :

$$\mathbb{P}(C) = \frac{W(z)\mathrm{e}^{-\underline{\mu}(z)}}{W^{\mu}(z)}.$$ □

4.4 Quelques calculs : spectre de fréquences allélique

Les expressions obtenues pour la population clonale de l'arbre nous permettent de calculer des espérances de quantités liées à toutes les mutations de l'arbre. En effet, si

f est une fonctionnelle des arbres réels (simples, munis de mutations et d'une mesure sur leurs feuilles), on peut chercher à connaître

$$\psi(z, f) := \mathbb{E}_z\left[\sum_{\substack{m\in\mathbb{T}\\ \text{mutation}}} f(\{\text{sous-arbre de } \mathbb{T} \text{ issu de } m\})\right],$$

où $\mathbb{T}$ sous la loi $\mathbb{P}_z$ est un CPP(ν, z), muni de mutations au taux μ. On peut calculer ces quantités, ce qui va nous permettre par exemple de mesurer l'intensité du spectre de fréquences allélique (défini dans ce qui suit). On définit, pour une mutation m sur l'arbre, l'ensemble R_m de la population qui porte m comme dernière mutation

$$R_m := \{t \in \mathbb{R}_+, \text{ la plus récente mutation sur la lignée de } t \text{ est } m\}.$$

Soit Φ la mesure ponctuelle aléatoire des tailles, mesurées avec la mesure de Lebesgue, des populations ayant le même allèle :

$$\Phi := \sum_{\substack{m\in\mathbb{T}\\ \text{mutation}}} \mathbf{1}_{R_m\neq\emptyset}\, \delta_{\mathrm{Leb}(R_m)}.$$

L'intensité Λ de ce processus ponctuel, c'est-à-dire la mesure sur $]0, \infty[$ telle que pour tout borélien B de $]0, \infty[$,

$$\Lambda(B) = \mathbb{E}_z[\Phi(B)],$$

est appelée l'intensité du spectre de fréquences allélique. L'analogue de cette mesure quand le nombre d'individu de la population finale est fini est la mesure moyenne $(\mathbb{E}A(k))_{k>0}$ du nombre $A(k)$ d'allèles portés exactement par k individus (notations $A_\theta(k, t)$ dans [6] et [3]). On cherche donc à identifier Λ, en remarquant que pour un borélien B,

$$\Lambda(B) = \psi(z, f_B),$$

avec $f_B(\mathbb{T}) := \mathbf{1}_{\mathrm{Leb}(R)\in B}$, où R désigne l'ensemble des feuilles clonales dans l'arbre ultramétrique $\mathbb{T}$. On obtient le résultat suivant :

Proposition 4.4. *Dans un CPP(ν, z) muni de mutations selon une mesure μ, sous les hypothèses* (H) *et avec les notations de la proposition 4.1, l'intensité du spectre de fréquences alléliques a la densité par rapport à la mesure de Lebesgue suivante :*

$$\frac{\Lambda(\mathrm{d}q)}{\mathrm{d}q} = W(z)\left(\frac{\mathrm{e}^{-\underline{\mu}(z)}}{W^\mu(z)^2}\mathrm{e}^{-q/W^\mu(z)} + \int_0^z \frac{\mathrm{e}^{-\underline{\mu}(x)}}{W^\mu(x)^2}\mathrm{e}^{-q/W^\mu(x)}\mu(\mathrm{d}x)\right).$$

Remarque 11. Cette expression est à comparer avec le corollaire 4.3 de [3] (le $(1-\frac{1}{W^\theta(x)})^{k-1}$ avec k discret devenant ici $\mathrm{e}^{-q/W^\mu(x)}$ avec q continu).

Remarque 12. En intégrant cette expression, on calcule l'espérance du nombre d'haplotypes différents dans la population :

$$\Lambda(\mathbb{R}_+) = \mathbb{E}_z[\Phi(\mathbb{R}_+)] = W(z)\left(\frac{\mathrm{e}^{-\underline{\mu}(z)}}{W^\mu(z)} + \int_0^z \frac{\mathrm{e}^{-\underline{\mu}(x)}}{W^\mu(x)}\mu(\mathrm{d}x)\right).$$

On remarque que $W(z)$ est la taille moyenne de la population totale de l'arbre, mesurée par la mesure de Lebesgue. Ainsi, le rapport du nombre d'haplotypes moyen sur le nombre d'individus moyen, dans la limite "grande population" ($z \to \infty$) est

$$\lim_{z\to\infty} \frac{\mathbb{E}_z[\Phi(\mathbb{R}_+)]}{W(z)} = \int_0^\infty \frac{\mathrm{e}^{-\underline{\mu}(x)}}{W^\mu(x)}\mu(\mathrm{d}x).$$

Toujours dans la limite "grande population", l'intensité du spectre devient

$$\lim_{z\to\infty} \frac{1}{W(z)}\left(\frac{\Lambda(\mathrm{d}q)}{\mathrm{d}q}\right) = \int_0^\infty \frac{\mathrm{e}^{-\underline{\mu}(x)}}{W^\mu(x)^2}\mathrm{e}^{-q/W^\mu(x)}\mu(\mathrm{d}x).$$

Cette expression est à comparer avec les limites de convergence presque sûre obtenues dans le théorème 3.1 de [6].

Démonstration. Ainsi, on cherche à calculer les $\psi(z, f)$, pour f une fonction mesurable positive d'un arbre réel muni de mutations et d'une mesure sur sa frontière. Supposons que l'on numérote les mutations $(M_n)_{n\geq 1}$ de $\mathbb{T}$ selon l'ordre des hauteurs (ordonnées) décroissantes. C'est-à-dire qu'on les numérote par ordre d'apparition dans le temps du processus de branchement. Alors on voit bien que pour tout $n \geq 1$, par la propriété de branchement, le sous-arbre issu de M_n a la loi de $\mathbb{T}$ sous $\mathbb{P}_{H_n}$, conditionnellement à la hauteur H_n de la mutation. Posons donc

$$\widetilde{f}(x) := \mathbb{E}_x[f(\mathbb{T})].$$

Alors l'expression que l'on cherche devient :

$$\begin{aligned}
\psi(z, f) &= \mathbb{E}_z\left[\sum_n f(\{\text{sous-arbre de } \mathbb{T} \text{ issu de } M_n\})\right] \\
&= \sum_n \mathbb{E}_z\left[f(\{\text{sous-arbre de } \mathbb{T} \text{ issu de } M_n\})\right] \\
&= \sum_n \mathbb{E}_z\left[\widetilde{f}(H_n)\right] \\
&= \mathbb{E}_z\left[\sum_n \widetilde{f}(H_n)\right].
\end{aligned}$$

Or cette expression est simple à calculer si l'on connaît $\widetilde{f}$ et l'intensité du processus ponctuel des hauteurs des mutations. En effet, par les propriétés des processus ponctuels :

$$\begin{aligned}
\mathbb{E}_z\left[\sum_n \widetilde{f}(H_n)\right] &= \mathbb{E}\left[\widetilde{f}(z) + \sum_{y\in M_{(0,z)}} \widetilde{f}(y) + \sum_{(t,x)\in\mathcal{N},\ t\leq T(z)}\left(\sum_{y\in M_{(t,x)}} \widetilde{f}(y)\right)\right] \\
&= \widetilde{f}(z) + \int_0^z \widetilde{f}(x)\mu(\mathrm{d}x) + \mathbb{E}\left[T(z)\int_0^z \nu(\mathrm{d}y)\int_0^y \widetilde{f}(x)\mu(\mathrm{d}x)\right] \\
&= \widetilde{f}(z) + \int_0^z \widetilde{f}(x)\mu(\mathrm{d}x) + \frac{1}{\overline{\nu}(z)}\int_0^z \widetilde{f}(x)(\overline{\nu}(x) - \overline{\nu}(z))\mu(\mathrm{d}x) \\
&= \widetilde{f}(z) + W(z)\int_0^z \frac{\widetilde{f}(x)}{W(x)}\mu(\mathrm{d}x).
\end{aligned}$$

On peut maintenant considérer $f(\mathbb{T})$ qui vaut 1 si et seulement si la taille de la population clonale est supérieure à q, pour un $q > 0$ fixé. Cela permet de calculer l'espérance du nombre de mutations qui affectent une population de taille supérieure à q, autrement dit $\Lambda(]q, \infty[)$. Comme on connaît la loi de la population clonale sous la probabilité $\mathbb{P}_z$, (voir le corollaire 4.2), on sait

$$\widetilde{f}(z) = \frac{W(z)\mathrm{e}^{-\underline{\mu}(z)}}{W^{\mu}(z)}\mathrm{e}^{-q/W^{\mu}(z)}.$$

Ainsi on a

$$\begin{aligned}\Lambda(]q, \infty[) &= \mathbb{E}_z[\Phi(]q, \infty[)] \\ &= W(z)\left(\frac{\mathrm{e}^{-\underline{\mu}(z)}}{W^{\mu}(z)}\mathrm{e}^{-q/W^{\mu}(z)} + \int_0^z \frac{\mathrm{e}^{-\underline{\mu}(x)}}{W^{\mu}(x)}\mathrm{e}^{-q/W^{\mu}(x)}\mu(\mathrm{d}x)\right).\end{aligned}$$

On n'a plus qu'à dériver une fois pour trouver l'expression de la proposition. □

A Annexes

A.1 Ensembles régénératifs, subordinateurs

Les subordinateurs sont les processus de Lévy croissants. Ils sont plutôt bien connus et engendrent naturellement des ensembles de réels aléatoires dits régénératifs. On va donc utiliser quelques propriétés des subordinateurs pour pouvoir caractériser les ensembles régénératifs.

Remarque 13. On notera $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (E, \mathcal{E}), (X_t)_{t\in\mathbb{R}_+}, (\mathbb{P}^{\mu})_{\mu\in\mathcal{M}^1(E)})$ pour désigner un processus à valeurs dans un espace mesurable $(E, \mathcal{E})$, défini sur un espace mesurable filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+})$, tel que sous la probabilité $\mathbb{P}^{\mu}$, le processus $(X_t)_{t\geq 0}$ est un processus de Markov adapté de loi initiale μ. On notera $\mathbb{P}^x$ pour $\mathbb{P}^{\delta_x}$ avec $x \in E$, et si l'on omet une partie de ces notations, on sous-entend que tous ces objets existent quand même. Aussi, quand on considère $\mathbb{R}$ en tant qu'espace, on sous-entend qu'il est muni de sa tribu borélienne. On peut toujours supposer que les tribus $\mathcal{F}$ et $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ sont complètes pour toutes les probabilités $(\mathbb{P}^{\mu})_{\mu\in\mathcal{M}^1(E)}$, et continues à droite.

Définition 7. Un subordinateur $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, (\sigma_t)_{t\in\mathbb{R}_+}, (\mathbb{P}^{\mu})_{\mu\in\mathcal{M}^1(\overline{\mathbb{R}})})$ est un processus de Markov fort à valeurs dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ à accroissements indépendants et stationnaires, croissant, continu à droite.

On va en général étudier un subordinateur $(\sigma_t)_{t\geq 0}$ sous la probabilité $\mathbb{P} := \mathbb{P}^0$, ce qui en fait un processus à valeurs dans $[0, \infty]$, issu de 0. Tout ce qu'il faut savoir sur les subordinateurs, en particulier les résultats suivants, se trouvent dans le cours de Saint-Flour de Bertoin [2].

Théorème A.1. *La loi d'un subordinateur $(\sigma_t)_{t\geq 0}$ est caractérisée par son exposant de Laplace ϕ, fonction croissante de $\mathbb{R}_+$ dans $\mathbb{R}_+$ telle que pour tous t, $\lambda > 0$,*

$$\mathbb{E}[\mathrm{e}^{-\lambda\sigma_t}] = \mathrm{e}^{-t\phi(\lambda)}.$$

On peut écrire ϕ sous la forme

$$\phi(\lambda) = d\lambda + k + \int_{]0,\infty[} (1 - \mathrm{e}^{-\lambda x})\pi(\mathrm{d}x),$$

où d est la dérive, k le taux de mort et π la mesure des sauts du subordinateur. On a $d \geq 0$, $k \geq 0$, et π vérifie

$$\int_{]0,\infty[} (x \wedge 1)\pi(\mathrm{d}x) < \infty.$$

Alors si $\zeta := \inf\{t \geq 0, \sigma_t = \infty\}$, on sait que ζ suit une loi exponentielle de paramètre k (si $k = 0$, alors $\zeta \equiv \infty$). Aussi on a presque sûrement pour tout $t < \zeta$,

$$\sigma_t = dt + \sum_{s \leq t} \Delta\sigma_s,$$

et l'ensemble des sauts $\{(s, \Delta\sigma_s), \Delta\sigma_s > 0\}$ forme un processus ponctuel de Poisson d'intensité $\mathrm{d}s \otimes \pi$.

On peut définir la mesure de renouvellement U d'un subordinateur, mesure sur $\mathbb{R}_+$, par

$$\int_{\mathbb{R}_+} f(x)U(\mathrm{d}x) = \mathbb{E}\left[\int_0^\infty f(\sigma_t)\mathrm{d}t\right].$$

Cette mesure caractérise la loi de $(\sigma_t)_{t\geq 0}$ puisque sa transformée de Laplace est l'inverse de ϕ :

$$\mathcal{L}U(\lambda) = \mathbb{E}\left[\int_0^\infty \mathrm{e}^{-\lambda\sigma_t}\mathrm{d}t\right] = \int_0^\infty \mathrm{e}^{-t\phi(\lambda)}\mathrm{d}t = \frac{1}{\phi(\lambda)}.$$

On remarque aussi qu'en posant $L_x = \inf\{t \geq 0, \sigma_t > x\}$ l'inverse continu à droite de (σ_t), on a

$$U(x) := U([0, x]) = \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\sigma_t \leq x}\mathrm{d}t\right] = \mathbb{E}[L_x].$$

A.1.1 Image d'un subordinateur

On considère un subordinateur $(\sigma_t)_{t\geq 0}$, et on note les objets associés comme dans la partie précédente. On pose $R = \overline{\{\sigma_t, t \geq 0\}} \subset \mathbb{R}_+$ la fermeture de son image. On définit une filtration $\mathcal{G}$ engendré par cet ensemble aléatoire par

$$\mathcal{G}_x := \bigcap_{y>x} \sigma(\mathbf{1}_{z\in R}, z \leq y).$$

Proposition A.2. *Cet ensemble R a la propriété régénérative suivante : pour tout $(\mathcal{G}_x)$-temps d'arrêt X qui prend ses valeurs dans l'ensemble des points de R non isolés à droite, sur l'événement $\{X < \infty\}$, on a l'égalité en loi*

$$R \cap [X, \infty[- X = \{x - X, x \in R, x \geq X\} \overset{(d)}{=} R,$$

et de plus $R \cap [X, \infty[- X$ est indépendant de $R \cap [0, X]$.

Démonstration. Quitte à remplacer X par $X\mathbf{1}_{\{X<\infty\}}$, on peut supposer que $X < \infty$ presque sûrement. Remarquons que $\mathcal{G}_x \subset \mathcal{F}_{L_x}$, où $\mathcal{F}$ est la filtration de départ, à laquelle est adapté σ, et L est l'inverse continu à droite de σ. Par continuité à droite de L, on sait que

$$\{L_X < t\} = \bigcup_{x\in\mathbb{Q}} (\{X < x\} \cap \{L_x < t\}).$$

Or $\{X < x\} \in \mathcal{F}_{L_x}$ donc pour tout x, on a $\{X < x\} \cap \{L_x < t\} \in \mathcal{F}_t$. D'où $\{L_X < t\} \in \mathcal{F}_t$, donc L_X est un $(\mathcal{F}_t)$-temps d'arrêt. Remarquons que comme presque sûrement $X \in R$ n'est pas isolé à droite, on a $\sigma_{L_X} = X$. On applique enfin la propriété de Markov forte : $(\sigma_{L_X+t} - X)_{t\geq 0}$ est un subordinateur de même loi que σ, issu de 0 et indépendant de $\mathcal{F}_{L_x}$, donc de $\mathcal{G}_x$, ce qui implique la propriété annoncée. □

A.1.2 Ensembles régénératifs

Cette propriété, appelée propriété régénérative, est en un sens propre aux images de subordinateurs. En effet, à quelques points isolés près, tous les ensembles aléatoires qui vérifient cette propriété peuvent être construits comme dans la partie précédente.

Soit un espace de probabilité filtré $(\Omega, \mathcal{G}, (\mathcal{G}_x)_{x\geq 0}, \mathbb{P})$ et un ensemble aléatoire $R \subset [0, \infty[$ progressivement mesurable, fermé, tel que $0 \in R$. On dit que R est régénératif si pour tout $(\mathcal{G}_x)$-temps d'arrêt X qui prend ses valeurs dans l'ensemble des points de R non isolés à droite, sur l'événement $\{X < \infty\}$, on a l'égalité en loi

$$R \cap [X, \infty[- X = \{x - X, x \in R, x \geq X\} \overset{(d)}{=} R,$$

et en plus $R \cap [X, \infty[- X$ est indépendant de $\mathcal{G}_X$.

Théorème A.3. *Soit R un ensemble régénératif sans point isolé. Alors il existe un subordinateur σ dont l'image est R.*

Démonstration. On prouve ce résultat seulement dans le cas le plus simple, si R a une mesure de Lebesgue non nulle. Plus précisément, on fait l'hypothèse $\mathbb{E}(\mathrm{Leb}(R)) > 0$. On peut poser pour $x \in \mathbb{R}_+$,

$$L_x := \mathrm{Leb}(R \cap [0, x]),$$

qui est clairement continue et son inverse σ. L'hypothèse que l'on fait assure que le temps d'arrêt

$$X := \inf\{x > 0, \mathrm{Leb}(R \cap [0, x]) > 0\}$$

n'est pas égal à $+\infty$ avec probabilité 1. Or, sur l'événement $\{X < \infty\}$, on sait que pour tout $\epsilon > 0$, on a $\mathrm{Leb}(R \cap [X, X + \epsilon]) > 0$. Donc X est un point de R non isolé à droite, et par la propriété de régénération,

$$(\mathrm{Leb}(R \cap [X, X + x]))_{x\geq 0} \overset{(d)}{=} (L_x)_{x\geq 0}.$$

En particulier, $X = 0$ presque sûrement a posteriori. Ceci assure que l'inverse continu à droite σ du processus L vérifie $\sigma_0 = 0$. On vérifie que σ est un subordinateur. En

effet, pour $t > 0$, on a σ_t est un $(\mathcal{G}_x)$-temps d'arrêt, et si $\sigma_t < \infty$, c'est que pour tout $\epsilon > 0$, on a $L_{\sigma_t+\epsilon} > t$. C'est-à-dire, pour tout $\epsilon > 0$,

$$\text{Leb}(R \cap [\sigma_t, \sigma_t + \epsilon) > 0.$$

Ceci implique que σ_t est un point de R non isolé à droite. Par régénération, l'ensemble des points de R à droite de σ_t est indépendant de $\mathcal{G}_{\sigma_t}$ et a même loi que R. Donc il est clair que $(\sigma_{t+s} - \sigma_t)_{s\geq 0}$ est indépendant de $\mathcal{G}_{\sigma_t}$ et a même loi que σ.

Il reste à montrer que σ est d'image dense dans R. Pour un rationnel positif q, on pose $D_q = \inf(R \cap [q, \infty[)$. C'est un temps d'arrêt, et comme R est supposé sans point isolé, D_q est non isolé à droite. Donc, en répétant l'argument déjà utilisé, (L_x) croît strictement au voisinage (à droite) de D_q donc $\sigma_{L_{D_q}} = D_q$. Ainsi on a les inclusions

$$\{D_q, q \in \mathbb{Q}_+\} \cap [0, \infty[\subset \{\sigma_t, t \in \mathbb{R}_+\} \subset R.$$

Or, comme $\{D_q, q \in \mathbb{Q}_+\} \cap [0, \infty[$ est dense dans R, R est bien la fermeture de l'image de σ.

Pour le cas général, voir la preuve complète dans Maisonneuve [7]. □

Dans ce cas particulier on l'on a écrit la preuve de la réciproque, on a en fait que $\mathbb{P}(x \in R)$ est la densité par rapport à la mesure de Lebesgue de la mesure de renouvellement de σ. En effet, on a :

$$\begin{aligned}
\int_{[0,\infty[} f(x)U(\mathrm{d}x) &= \mathbb{E}\left[\int_0^\infty f(\sigma_t)\mathrm{d}t\right] \\
&= \mathbb{E}\left[\int_{[0,\infty[} f(x)\mathrm{d}L_x\right] \\
&= \mathbb{E}\left[\int_0^\infty f(x)\mathbf{1}_{x\in R}\mathrm{d}x\right] \\
&= \int_0^\infty f(x)\mathbb{P}(x \in R)\mathrm{d}x.
\end{aligned}$$

On peut aussi remarquer que le terme de dérive de σ ainsi construit est toujours 1. D'une manière générale, si σ'_t est un subordinateur de dérive d, on a

$$dt = \sigma'_t - \sum_{s\leq t} \Delta\sigma'_s = \text{Leb}([0, \sigma'_t] \cap R'),$$

où R' désigne l'image de σ'. Or on a construit σ à partir de R tel que

$$\text{Leb}([0, \sigma_t] \cap R) = L_{\sigma_t} = t,$$

donc le terme de dérive de σ est $d = 1$.

A.2 Mesures ponctuelles de Poisson

Soit $\mathcal{N}$ une mesure ponctuelle de Poisson sur un espace mesuré $(E, \mathcal{E})$ d'intensité μ mesure σ-finie sur $\mathcal{E}$. On rappelle sans démonstration des résultats que l'on trouve dans un cours de M2 sur les processus ponctuels.

Proposition A.4. *Pour toute fonction mesurable $f : E \to \mathbb{R}_+$, on a*

$$\mathbb{E}\left[\sum_{x\in\mathcal{N}} f(x)\right] = \int_E f \mathrm{d}\mu.$$

De plus, on a l'alternative suivante.

$$\int_E f(x)\wedge 1\,\mu(\mathrm{d}x) < \infty \quad \Longrightarrow \quad \sum_{x\in\mathcal{N}} f(x) < \infty \quad p.s,$$
$$\int_E f(x)\wedge 1\,\mu(\mathrm{d}x) = \infty \quad \Longrightarrow \quad \sum_{x\in\mathcal{N}} f(x) = \infty \quad p.s.$$

On connaît également la transformée de Laplace du processus :

$$\mathbb{E}\left[\mathrm{e}^{-\sum_{X\in\mathcal{N}} f(X)}\right] = \exp\left(-\int_E (1-\mathrm{e}^{-f(x)})\mu(\mathrm{d}x)\right).$$

A.2.1 Branchement

Soit $\mathcal{N}$ une mesure ponctuelle de Poisson sur $\mathbb{R}_+^2$ d'intensité $\mathrm{d}t\otimes\nu$, avec ν mesure diffuse sur $\mathbb{R}_+$, telle que pour tout $z > 0$, $\overline{\nu}(z) < \infty$. On suppose aussi pour simplifier que $\overline{\nu}(0) = \infty$, c'est-à-dire que ν est de masse infinie. Soit $z > 0$ tel que $\overline{\nu}(z) > 0$. On appelle $\mathcal{N}_z$ l'ensemble de points de $\mathcal{N}$ avant le premier point de hauteur plus grande que z :

$$\mathcal{N}_z := \mathcal{N} \cap ([0, T(z)[\times[0, z]),$$

où $T(z) = \inf\{t \geq 0, (x, t) \in \mathcal{N}, x \geq z\}$ est la largeur de $\mathcal{N}_z$. On pose $\mathcal{P}^z$ la loi du couple $(\mathcal{N}_z, T(z))$. Les ensembles de points aléatoires $(\mathcal{N}_z, T(z))$ associés à leur largeur vérifient la propriété de branchement suivante.

Soit $Z := \max\{x \geq 0, (t, x) \in \mathcal{N}_z\}$ la hauteur du point le plus haut dans $\mathcal{N}_z$, avec $Z = 0$ ssi $\mathcal{N}_z = \emptyset$. Conditionnellement à Z, on se donne des variables aléatoires $(\mathcal{N}_1, T_1)$ et $(\mathcal{N}_2, T_2)$ indépendantes et de même loi $\mathcal{P}^Z$. On recolle $\mathcal{N}_1$ et $\mathcal{N}_2$ côte à côte, en rajoutant un point de hauteur Z entre les deux :

$$\widetilde{\mathcal{N}} = \mathcal{N}_1 \cup \{(T_1, Z)\} \cup \{(T_1 + t, x), (t, x) \in \mathcal{N}_2\}.$$

Proposition A.5. *Sous les hypothèses et avec les notations précédentes, on a l'égalité en loi*

$$(\widetilde{\mathcal{N}}, T_1 + T_2) \overset{(d)}{=} (\mathcal{N}_z, T(z)).$$

Démonstration. On connaît la loi de $(\mathcal{N}_z, T(z))$: $T(z)$ est une variable exponentielle de paramètre $\overline{\nu}(z)$ et conditionnellement à $T(z)$, $\mathcal{N}_z$ est une mesure ponctuelle de Poisson sur $[0, T(z)[\times[0, z]$ d'intensité $\mathrm{d}t \times \nu$. Toujours conditionnellement à $T(z)$, le plus haut atome de $\mathcal{N}_z$ est (U, Z), avec U uniforme sur $[0, T(z)]$ et Z indépendant de U, tel que

$$\mathcal{P}^z(Z \leq x \mid T(z)) = \mathrm{e}^{-T(z)(\overline{\nu}(x)-\overline{\nu}(z))}.$$

La loi jointe de $(Z, T(z))$ est donc donné par :

$$\begin{aligned}
\mathbb{E}[f(T(z))\mathbf{1}_{Z\leq x}] &= \int_0^\infty \overline{\nu}(z)\mathrm{e}^{-\overline{\nu}(z)t}\mathrm{e}^{-T(z)(\overline{\nu}(x)-\overline{\nu}(z))}\mathrm{d}t \\
&= \int_0^\infty \overline{\nu}(z)\mathrm{e}^{-\overline{\nu}(x)t}\mathrm{d}t \\
&= \int_0^\infty \overline{\nu}(z)\int_{\overline{\nu}(x)}^\infty t\mathrm{e}^{-ut}\mathrm{d}u\mathrm{d}t \\
&= \int_{\overline{\nu}(x)}^\infty \frac{\overline{\nu}(z)}{u^2}\int_0^\infty tu^2\mathrm{e}^{-ut}\mathrm{d}t\mathrm{d}u
\end{aligned}$$

Autrement dit, vu nos hypothèses sur ν, la variable aléatoire $\overline{\nu}(Z)$ a une densité $\frac{\overline{\nu}(z)}{u^2}\mathbf{1}_{u\geq\overline{\nu}(z)}\mathrm{d}u$, et conditionnellement à $\overline{\nu}(Z)$, $T(z)$ est une loi gamma de paramètre $(\overline{\nu}(Z), 2)$. Comme $U/T(z)$ est uniforme sur $[0, 1]$ et indépendante de Z, on peut vérifier que $(Z, T(z), U)$ a bien la même loi que $(Z, T_1 + T_2, T_1)$, où conditionnellement à Z, T_1 et T_2 sont indépendantes de loi exponentielle de paramètre $\overline{\nu}(Z)$. □

A.3 Divers

Lemme A.6 (Dérivation de fonctions composées)**.** *Soit $F : \mathbb{R} \to \mathbb{R}$ continue monotone, et $\phi : \mathbb{R} \to \mathbb{R}$ dérivable monotone. Alors on a l'égalité suivante*

$$\mathrm{d}\phi \circ F(x) = \phi'(F(x))\mathrm{d}F(x)$$

Démonstration. Quitte à faire quelques remplacements inoffensifs, on suppose que F et ϕ sont toutes les deux croissantes. On introduit les inverses continus à droite de ces fonctions F^{-1} et ϕ^{-1}, et on remarque les faits suivants, pour tous $a < b$ et pour toute fonction f borélienne positive,

$$\begin{aligned}
F(a) \leq x < F(b) &\iff a \leq F^{-1}(x) < b, \\
\text{et } \int_{\mathbb{R}} f(F^{-1}(x))\mathrm{d}x &= \int_{\mathbb{R}} f(x)\mathrm{d}F(x).
\end{aligned}$$

Aussi par continuité, on a pour tout x dans l'image de F, $x = F(F^{-1}(x))$. On peut finalement écrire pour tous $a < b$,

$$\begin{aligned}
\int_a^b \mathrm{d}\phi \circ F(x) = \phi(F(b)) - \phi(F(a)) &= \int_{\mathbb{R}} \mathbf{1}_{\{\phi(F(a))\leq t<\phi(F(b))\}}\mathrm{d}t \\
&= \int_{\mathbb{R}} \mathbf{1}_{\{F(a)\leq\phi^{-1}(t)<F(b)\}}\mathrm{d}t \\
&= \int_{\mathbb{R}} \mathbf{1}_{\{F(a)\leq u<F(b)\}}\phi'(u)\mathrm{d}u \\
&= \int_{\mathbb{R}} \mathbf{1}_{\{a\leq F^{-1}(u)<b\}}\phi'(F(F^{-1}(u)))\mathrm{d}u \\
&= \int_a^b \phi'(F(v))\mathrm{d}F(v),
\end{aligned}$$

D'où l'égalité des mesures. □

Références

[1] R. Abraham et L. Serlet. Poisson Snake and Fragmentation. *Electronic Journal of Probability*, 7 (2002). paper no. 17. DOI : 10.1214/EJP.v7-116 (cf. p. 21).

[2] J. Bertoin. Subordinators : Examples and Applications. In : *Lectures on Probability Theory and Statistics : École d'Été de Probabilités de Saint-Flour XXVII*. Springer, 1997, p. 1–91. DOI : 10.1007/978-3-540-48115-7_1 (cf. p. 25).

[3] N. Champagnat et A. Lambert. Splitting Trees with Neutral Poissonian Mutations I : Small Families. *Stochastic Processes and their Applications*, 122.3 (2012), p. 1003–1033. DOI : 10.1016/j.spa.2011.11.002 (cf. p. 2, 8, 23).

[4] W. J. Ewens. The Sampling Theory of Selectively Neutral Alleles. *Theoretical Population Biology*, 3.1 (1er mar. 1972), p. 87–112. DOI : 10.1016/0040-5809(72)90035-4 (cf. p. 2, 8).

[5] A. Lambert et G. Uribe Bravo. The Comb Representation of Compact Ultrametric Spaces. *p-Adic Numbers, Ultrametric Analysis and Applications*, 9.1 (jan. 2017), p. 22–38. DOI : 10.1134/S2070046617010034 (cf. p. 6).

[6] A. Lambert. The Allelic Partition for Coalescent Point Processes. *Markov Processes and Related Fields*, 15 (2009), p. 359–386. arXiv : 0804.2572 (cf. p. 17, 23, 24).

[7] B. Maisonneuve. Ensembles régénératifs, temps locaux et subordinateurs. In : *Séminaire de Probabilités V Université de Strasbourg*. Lecture Notes in Mathematics 191. Springer Berlin Heidelberg, 1971, p. 147–169. DOI : 10.1007/BFb0058856 (cf. p. 28).

[8] P. Marchal. Nested Regenerative Sets and Their Associated Fragmentation Process. In : *Mathematics and Computer Science III*. Trends in Mathematics. Birkhäuser Basel, 2004, p. 461–470. DOI : 10.1007/978-3-0348-7915-6_45 (cf. p. 9, 19).