Arbres de fragmentation

Adrien Joseph

20 mai 2008

1 Introduction

On commence par définir la notion de fragmentation en temps discret. On dispose initialement d'une masse $m > 0$. À la date $t = 1$, on la fragmente aléatoirement en $m_1, m_2, \dots$, c'est-à-dire que pour tout $i \geq 1$, $m_i \geq 0$ et $\sum_i m_i \leq m$. On suppose que la suite (m_i) est décroissante. À la date $t = 2$, les masses m_i se fragmentent : pour tout i, il existe une suite décroissante $(m_{i,j})_{j\geq 1}$ de réels positifs dont la somme est inférieure ou égale à m_i. On suppose que les masses m_i se fragmentent indépendamment les unes des autres et qu'on a invariance d'échelle, i.e. pour tout i tel que $m_i > 0$,

$$\left(\frac{m_1}{m}, \frac{m_2}{m}, \dots\right) \overset{\text{(loi)}}{=} \left(\frac{m_{i,1}}{m_i}, \frac{m_{i,2}}{m_i}, \dots\right).$$

On réitére à chaque instant entier ce procédé : chaque fragment se casse aléatoirement indépendamment des autres fragments et en suivant la même loi. On voit ainsi que la loi d'une telle fragmentation est entièrement déterminée par la distribution de la suite $(m_1/m, m_2/m, \dots)$. On souhaite définir maintenant des fragmentations en temps continu.

On considère un objet qui se casse au cours du temps. On suppose que le processus de fragmentation, qui donne les masses des fragments en fonction du temps, vérifie les conditions suivantes. Connaissant la fragmentation à l'instant s, la fragmentation à la date t est obtenue en cassant les masses présentes à la date s indépendamment les unes des autres. On suppose que la variable aléatoire donnant le rapport des masses des fragments obtenus à la date t ne dépend que l'accroissement $t - s$: la suite décroissante des masses des fragments présents à l'instant t provenant d'un unique objet de masse $m > 0$ présent à la date s a la même loi que $mF(t - s)$, où $F(u)$ est la suite décroissante des masses des fragments présents à la date u issus de la fragmentation d'un objet initialement de masse 1. Un tel processus est appelé *fragmentation homogène*.

L'idée fondamentale est la suivante : on considère une suite de points aléatoires U_i i.i.d. suivant une loi adaptée à la distribution de la masse de l'objet. On suppose que ces v.a. sont indépendantes du processus de fragmentation. On définit alors un processus $\Pi(\cdot)$ à valeurs dans l'ensemble des partitions de $\mathbf{N}$ en déclarant qu'à la date t, deux entiers i et j appartiennent à un même bloc de la partition $\Pi(t)$ si et seulement si les points U_i et U_j appartiennent à un même fragment de l'objet à la date t. On remarque que le processus $\Pi(\cdot)$ se raffine, i.e. si deux entiers appartiennent à un même bloc de $\Pi(t)$, alors ils appartiennent à un même bloc de $\Pi(s)$ pour tout $s \leq t$. Le processus $\Pi(\cdot)$ est aussi *échangeable*, i.e. une permutation d'indices n'affecte pas sa loi. La loi des grands nombres assure que les masses des fragments à un instant t peuvent être connues : ce sont les fréquences asymptotiques des blocs de la partition $\Pi(t)$. On s'intéresse donc uniquement aux processus de fragmentations homogènes à valeurs dans les partitions de $\mathbf{N}$.

2 Partitions de N

2.1 Définitions

Soit B une partie de $\mathbf{N}$. Une partition $\pi = (\pi_1, \pi_2, \dots)$ de B est une suite infinie de blocs $\pi_i \subset B$ disjoints deux à deux et dont la réunion est B. On convient de plus que les blocs sont ordonnés de la manière suivante : si $i \leq j$, alors $\min \pi_i \leq \min \pi_j$, avec la convention $\min \emptyset = \infty$. Si $i \in B$, le bloc de la partition π contenant i est noté $\pi_{(i)}$. On note $\mathcal{P}_B$ l'ensemble des partitions de B. Pour tout $n \in \mathbf{N}$, l'ensemble $\mathcal{P}_{\{1,\dots,n\}}$ est noté $\mathcal{P}_n$ et $\mathcal{P}_{\mathbf{N}}$ simplement $\mathcal{P}$. La partition $(B, \emptyset, \dots)$ est dite triviale et est notée $\mathbf{1}_B$. Si $B = \{i_1, i_2, \dots\}$, la partition $(\{i_1\}, \{i_2\}, \dots)$ composée uniquement de singletons est notée $\mathbf{0}_B$ et est appelée la partition discrète de B.

On peut voir une partition π de B comme une relation d'équivalence en disant que deux éléments i et j sont équivalents (on note $i \overset{\pi}{\sim} j$) si et seulement s'ils appartiennent au même bloc de π. Pour toutes partitions π et π' d'un ensemble B, on dit que π est plus fine que π' si pour tous $i, j \in B$, $i \overset{\pi}{\sim} j$ implique $i \overset{\pi'}{\sim} j$. Pour toute permutation $\sigma : \mathbf{N} \to \mathbf{N}$, la relation

$$i \sim j \Longleftrightarrow \sigma(i) \overset{\pi}{\sim} \sigma(j), \quad i, j \in \mathbf{N}$$

est une relation d'équivalence, et la partition de $\mathbf{N}$ associée est notée $\sigma(\pi)$. On dit que la permutation σ est *finie* si $\sigma(n) = n$ pour n assez grand.

Si B' est une partie de B et π une partition de B, alors la relation

$$i \sim j \Longleftrightarrow i \overset{\pi}{\sim} j, \quad i, j \in B'$$

est une relation d'équivalence. La partition de B' associée est notée $\pi \circ B'$ et est appelée partition π restreinte à B'. Pour tout $\pi \in \mathcal{P}$, on note $\pi_{|n} = \pi \circ \{1, \dots, n\}$. Remarquons qu'on a la *propriété de consistance* suivante : si $n < n'$, la partition $\pi_{|n}$ est la restriction de $\pi_{|n'}$ à $\{1, \dots, n\}$. Réciproquement, si la suite de partitions γ_n de $\{1, \dots, n\}$ est consistante, alors il existe une partition $\pi \in \mathcal{P}$ telle que $\pi_{|n} = \gamma_n$.

On munit l'ensemble $\mathcal{P}$ de la distance $\Delta(\pi, \pi') = 2^{-N(\pi, \pi')}$, où $N(\pi, \pi')$ est le plus grand entier n tel que $\pi_{|n} = \pi'_{|n}$. Alors $(\mathcal{P}, \Delta)$ est un espace métrique compact. On le munit de sa tribu borélienne.

2.2 Fragmentations homogènes

Commençons par définir la notion de continuité en probabilité.

Définition 1 *Un processus $(X_t, t \geq 0)$ à valeurs dans un espace métrique est dit continu en probabilité si pour tout $t \geq 0$, X_s converge en probabilité vers X_t lorsque s tend vers t.*

On considère un processus $(\Pi(t), t \geq 0)$ à valeurs dans $\mathcal{P}$ continu en probabilité partant p.s. de la partition triviale. Le processus $(\Pi(t), t \geq 0)$ est dit *échangeable* si pour toute permutation finie σ, le processus $(\sigma(\Pi(t)), t \geq 0)$ a même loi que $\Pi(\cdot)$.

Définition 2 *Le processus échangeable* $(\Pi(t), t \geq 0)$ *est dit de fragmentations homogènes si presque sûrement il n'est pas constant et si pour tous* $t_0, t > 0$, *conditionnellement à* $\Pi(t_0) = (\pi_1, \pi_2, \dots)$, *la partition* $\Pi(t_0 + t)$ *est indépendante de* $(\Pi(s), 0 \leq s \leq t_0)$ *et a même loi que la partition de* $\mathbf{N}$ *dont les blocs sont* $\gamma^{(1)} \circ \pi_1, \gamma^{(2)} \circ \pi_2, \dots$, *où les* $\gamma^{(i)}$ *sont des v.a. indépendantes suivant la loi de* $\Pi(t)$.

Remarquons que pour tous $0 \leq t \leq t'$, la partition $\Pi(t')$ est plus fine que la partition $\Pi(t)$. Cela nous permet de choisir une version continue à droite ayant une limite à gauche en tout point (càdlàg) de $\Pi(\cdot)$: on définit pour tout $t \geq 0$ la partition $\Pi(t+)$ associée à la relation d'équivalence suivante :

$$i \overset{\Pi(t+)}{\sim} j \Longleftrightarrow i \overset{\Pi(t')}{\sim} j \text{ pour un rationnel } t' > t,$$

et par l'hypothèse de continuité en probabilité, $(\Pi(t+), t \geq 0)$ est une version càdlàg de $\Pi(\cdot)$; on considère à partir de maintenant uniquement cette version. Notons aussi que $\Pi(\infty-)$ est la partition discrète.

Un processus de fragmentations homogènes est markovien, et son semi-groupe vérifie la propriété de Feller. Pour plus de détails, on renvoie à [1].

3 Mesures échangeables de partitions

3.1 Mesure caractéristique

Soit $\Pi(\cdot)$ un processus de fragmentations homogènes. Pour tout $n \geq 2$, on note $\Pi_{|n}(\cdot)$ le processus restreint à $\{1, \dots, n\}$. Alors $\Pi_{|n}(\cdot)$ est une chaîne de Markov à temps continu. La date T_n à laquelle $\Pi_{|n}(\cdot)$ quitte son état initial (la partition triviale de $\{1, \dots, n\}$) suit une loi exponentielle de paramètre $q(n) \in]0, \infty[$. On note $\rho(n)$ la loi de $\Pi_{|n}(T_n)$. D'après la définition d'un processus de fragmentations homogènes, on voit que la suite $(q(n), \rho(n))_{n \in \mathbf{N}}$ caractérise la loi de $\Pi(\cdot)$.

Considérons maintenant $\Pi_{|n+1}(T(n+1))$: ou bien $T(n+1) < T(n)$, et c'est la partition dont les seuls blocs non vides sont $\{1, \dots, n\}$ et $\{n+1\}$, ou $T(n+1) = T(n)$ et sa restriction à $\{1, \dots, n\}$ n'est pas la partition triviale. Par conséquent, le taux de saut de $\Pi_{|n+1}(\cdot)$ sur l'ensemble des partitions de $\{1, \dots, n+1\}$ dont la restriction à $\{1, \dots, n\}$ n'est pas triviale est $q(n)$, et la loi de la restriction de $\Pi_{|n+1}(T_{n+1})$ à $\{1, \dots, n\}$ conditionnellement à l'événement $\{T(n+1) = T(n)\}$ est $\rho(n)$ (on obtient de nouveau une propriété de consistance). En notant ϕ_n la restriction de l'application canonique $\mathcal{P}_{n+1} \to \mathcal{P}_n$ à l'ensemble des partitions de $\{1, \dots, n+1\}$ non triviales sur $\{1, \dots, n\}$, on en déduit que la mesure image de $q(n+1)\rho(n+1)$ par ϕ_n est $q(n)\rho(n)$.

Notons $\mathcal{P}_n^*$ l'ensemble des partitions $\pi \in \mathcal{P}$ telles que la restriction de π à $\{1, \dots, n\}$ est non triviale. Par le théorème d'extension de Kolmogorov, il existe une unique mesure κ sur $\mathcal{P}$ qui ne donne aucune masse à la partition triviale et telle que pour tout $n \in \mathbf{N}$, la mesure image de la restriction de κ à $\mathcal{P}_n^*$ par l'application canonique $\mathcal{P}_n^* \to \mathcal{P}_n$ est

$q(n)\rho(n)$. En particulier, $\kappa(\mathcal{P}_n^*) = q(n) < \infty$. En outre, comme pour tout $n \in \mathbf{N}$, $\rho(n)$ est invariante sous l'action des permutations de $\{1, \ldots, n\}$ par le caractère échangeable de $\Pi(\cdot)$, κ est invariante sous l'action des permutations finies. Introduisons la définition suivante :

Définition 3 *Une mesure κ sur $\mathcal{P}$ est une mesure échangeable de partitions si :*

1. *κ est échangeable, i.e. invariante sous l'action des permutations finies,*
2. *κ ne charge pas la partition triviale,*
3. *$\kappa(\mathcal{P}_n^*) < \infty$ pour tout $n \in \mathbf{N}$.*

Puisque les ensembles $\mathcal{P}_n^*$, $n \geq 1$, sont compacts et que le complémentaire de $\bigcup_{n\in\mathbf{N}} \uparrow \mathcal{P}_n^*$ est réduit à la partition triviale, une mesure échangeable de partitions est σ-finie.

On déduit de nos remarques précédentes la proposition suivante :

Proposition 1 *Soit $\Pi(\cdot)$ un processus de fragmentations homogènes. Il existe une unique mesure échangeable de partitions κ telle que pour tout $n \in \mathbf{N}$ et pour toute partition non triviale γ_n de $\mathcal{P}_n$, on a :*

$$\kappa(\{\pi \in \mathcal{P}, \pi_{|n} = \gamma_n\}) = \frac{1}{\mathbf{E}[T_n]} \mathbf{P}(\Pi_{|n}(T_n) = \gamma_n),$$

où $\Pi_{|n}(\cdot)$ est le processus restreint à $\{1, \ldots, n\}$ et $T_n = \inf\{t \geq 0, \Pi(t) \in \mathcal{P}_n^*\}$. En outre, κ détermine entièrement la loi de $\Pi(\cdot)$; on l'appelle la mesure caractéristique de $\Pi(\cdot)$.

3.2 Construction d'un processus de fragmentations homogènes

Montrons que réciproquement, en partant d'une mesure échangeable de partitions κ, on peut obtenir un processus de fragmentations homogènes dont la mesure caractéristique est κ. On suit la construction proposée par Bertoin dans [1].

Introduisons tout d'abord une nouvelle notation. Pour toute partition $\pi = (\pi_1, \pi_2, \ldots)$ d'une partie B de $\mathbf{N}$, pour tout entier k et pour toute partition Δ de $\mathbf{N}$, on note

$$\Delta \overset{k}{\circ} \pi \in \mathcal{P}_B$$

la partition de B dont les blocs sont les π_i, $i \neq k$, et les blocs de la partition $\Delta \circ \pi_k$ de π_k.

Comme la mesure κ est σ-finie, il existe un processus ponctuel de Poisson temporel $((\Delta_t, k_t), t \geq 0)$ sur $\mathcal{P} \times \mathbf{N}$ d'intensité $\kappa \otimes \sharp$, où $\sharp$ est la mesure de comptage sur $\mathbf{N}$. Puisque $\kappa(\mathcal{P}_n^*) < \infty$, le processus de Poisson restreint à $\mathcal{P}_n^* \times \{1, \ldots, j\}$ est p.s. discret pour tout entier j. On va montrer qu'il existe un processus $\Pi(\cdot)$ càdlàg à valeurs dans $\mathcal{P}$ tel que, si le processus de Poisson n'a pas de point à l'instant t, $\Pi(t) = \Pi(t-)$, et sinon :

$$\Pi(t) = \Delta_t \overset{k_t}{\circ} \Pi(t-). \tag{1}$$

Soit $n \in \mathbf{N}$. Notons que pour toute partition π de $\{1, \ldots, n\}$, pour tout $k \in \mathbf{N}$ et pour tout $\Delta \in \mathcal{P}$, si la restriction de Δ au k-ème bloc de π est triviale, alors $\Delta \overset{k}{\circ} \pi = \pi$. On sait aussi que l'ensemble des instants t où $\Delta_t \in \mathcal{P}_n^*$ et $k_t \leq n$ est p.s. discret. On peut donc construire un processus càdlàg $\pi^{(n)}(\cdot)$ à valeurs dans $\mathcal{P}_n$ partant de la partition triviale tel que $\pi^{(n)}(\cdot)$ saute seulement aux instants t correspondant aux points (Δ_t, k_t), et dans ce cas,

$$\pi^{(n)}(t) = \Delta_t \overset{k_t}{\circ} \pi^{(n)}(t-).$$

Remarquons en effet qu'un tel instant t induit un changement dans le processus seulement lorsque $\Delta_t \in \mathcal{P}_n^*$ et $k_t \leq n$.

Il est clair que pour tous $k, n \in \mathbf{N}$, $\Delta, \pi \in \mathcal{P}$,

$$\left(\Delta \overset{k}{\circ} \pi\right)_{|n} = \Delta \overset{k}{\circ} \pi_{|n}.$$

La famille $(\pi^{(n)}(\cdot), n \in \mathbf{N})$ vérifie donc la propriété de consistance. On en déduit l'existence d'un processus càdlàg $\Pi(\cdot)$ à valeurs dans $\mathcal{P}$ dont la restriction $\Pi_{|n}(\cdot)$ à $\{1, \ldots, n\}$ coïncide avec $\pi^{(n)}(\cdot)$ pour tout $n \in \mathbf{N}$. Le processus $\Pi(\cdot)$ satisfait (1).

On peut montrer le résultat fondamental suivant :

Théorème 1 *Le processus $\Pi(\cdot)$ est de fragmentations homogènes de mesure. Sa mesure caractéristique est κ.*

3.3 Caractérisation des fragmentations

On souhaite maintenant étudier les mesures échangeables de partitions, qui sont d'après ce qu'on vient de voir en correspondance bijective avec les lois des processus de fragmentations homogènes. Donnons tout d'abord des exemples de telles mesures.

Pour tout $n \in \mathbf{N}$, on note ε_n la partition de $\mathbf{N}$ qui a pour blocs non vides $\{n\}$ et $\mathbf{N}\backslash\{n\}$. On note pour toute partition $\pi \in \mathcal{P}$, δ_π la mesure de Dirac en π. Posons pour tout $c \geq 0$,

$$\mu_c = c \sum_{n=1}^{\infty} \delta_{\varepsilon_n}.$$

Il est clair que la mesure μ_c est une mesure échangeable de partitions. On appelle μ_c la *mesure d'érosion* de taux d'érosion c.

Un autre exemple est fourni en introduisant les ensembles suivants :

$$\begin{aligned} \mathcal{S}^{\downarrow} &= \{s = (s_i, i \in \mathbf{N}), s_1 \geq s_2 \geq \cdots \geq 0 \text{ et } \textstyle\sum_{i=1}^{\infty} s_i \leq 1\}, \\ \mathcal{S}^* &= \mathcal{S}^{\downarrow}\backslash\{(1, 0, \ldots)\}. \end{aligned}$$

Soit $s \in \mathcal{S}^*$. Reprenant la construction de Kingman dans [4], on associe à s une partition aléatoire échangeable. On considère tout d'abord une suite X_i, $i \in \mathbf{N}$, de variables aléatoires i.i.d. à valeurs entières telles que :

$$\mathbf{P}(X_i = n) = s_n \text{ si } n \in \mathbf{N} \text{ et } \mathbf{P}(X_i = 0) = 1 - \sum_{n=1}^{\infty} s_n.$$

On définit alors une partition aléatoire de $\mathbf{N}$ associée à la relation d'équivalence suivante :

$$i \sim j \Longleftrightarrow i = j \text{ ou } X_i = X_j > 0.$$

Remarquons que par indépendance et équidistribution des X_i, cette partition aléatoire est échangeable ; on note μ_s sa loi. On peut montrer que l'application $s \mapsto \mu_s$ est continue. L'idée est de définir une mesure sur $\mathcal{P}$ en mélangeant les mesures μ_s via une mesure sur $\mathcal{S}^*$. Donnons tout d'abord la définition suivante :

Définition 4 *On appelle mesure de Lévy sur $\mathcal{S}^*$ toute mesure σ-finie ν sur $\mathcal{S}^*$ telle que, en notant $s_1 < 1$ le premier terme de la suite $s \in \mathcal{S}^*$,*

$$\int_{\mathcal{S}^*} (1 - s_1)\nu(ds) < \infty.$$

La proposition suivante fournit un second exemple de mesures échangeables de partitions :

Proposition 2 *Soit ν une mesure de Lévy sur $\mathcal{S}^*$. On définit la mesure μ_ν sur $\mathcal{P}$ par :*

$$\mu_\nu(d\pi) = \int_{\mathcal{S}^*} \mu_s(d\pi)\nu(ds).$$

Alors μ_ν est une mesure échangeable de partitions.

Remarque On va voir dans le théorème 2 que la connaissance de μ_ν permet de retrouver la mesure ν. On appelle μ_ν la *mesure de dislocation* de mesure de Lévy ν.

Le résultat principal de cette section assure que toute mesure échangeable de partitions κ se décompose canoniquement en une somme d'une mesure d'érosion μ_c et d'une mesure de dislocation μ_ν : $\kappa = \mu_c + \mu_\nu$. On commence par la définition suivante :

Définition 5 *Soit $\pi = (\pi_1, \pi_2, \dots) \in \mathcal{P}$. On dit que la partition π a des fréquences asymptotiques si pour tout $i \in \mathbf{N}$, la limite suivante existe :*

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{card}\{j \le n, j \in \pi_i\} = \lambda_i \in [0, 1].$$

Si on ordonne les λ_i en $\lambda_1^\downarrow \ge \lambda_2^\downarrow \ge \dots$, alors, en posant $\Lambda(\pi) = (\lambda_1^\downarrow, \lambda_2^\downarrow, \dots)$, on a $\Lambda(\pi) \in \mathcal{S}^\downarrow$.

On peut maintenant citer le résultat suivant, prouvé dans [1].

Théorème 2 *Soit κ une mesure échangeable de partitions. Alors il existe un unique réel $c \ge 0$ et une unique mesure de Lévy ν sur $\mathcal{S}^*$ tels que $\kappa = \mu_c + \mu_\nu$. En outre :*

1. *pour κ-presque tout $\pi \in \mathcal{P}$, π a des fréquences asymptotiques $\Lambda(\pi)$,*
2. *la restriction de κ à l'ensemble des partitions π telles que $\Lambda(\pi) \neq (1, 0, \dots)$ est une mesure de dislocation : en notant $\Lambda(\kappa)$ la mesure image de κ par l'application mesurable $\pi \mapsto \Lambda(\pi)$, la restriction $\nu = \mathbf{1}_{\mathcal{S}^*}\Lambda(\kappa)$ de $\Lambda(\kappa)$ à $\mathcal{S}^*$ est une mesure de Lévy sur $\mathcal{S}^*$ et*
$$\mathbf{1}_{\{\Lambda(\pi)\in\mathcal{S}^*\}}\kappa(d\pi) = \mu_\nu(d\pi),$$
3. *la restriction de κ à l'ensemble des partitions π telles que $\Lambda(\pi) = (1, 0, \dots)$ est une mesure d'érosion : il existe $c \ge 0$ tel que*
$$\mathbf{1}_{\{\Lambda(\pi)=(1,0,\dots)\}}\kappa(d\pi) = \mu_c(d\pi).$$

4 Fragment marqué

Soit $\Pi(\cdot)$ un processus de fragmentations homogènes de mesure caractéristique κ se décomposant canoniquement en $\kappa = \mu_c + \mu_\nu$. On souhaite étudier l'évolution du bloc de la partition $\Pi(t)$ contenant un élément donné, disons 1, au cours du temps. Plus précisément, notons, pour tout $t \geq 0$, $\overline{\lambda}_1(t)$ et $\underline{\lambda}_1(t)$ respectivement les fréquences asymptotiques supérieure et inférieure du premier bloc de $\Pi(t)$:

$$\begin{array}{rcl} \overline{\lambda}_1(t) & = & \limsup_{n\to\infty} n^{-1}\,\mathrm{card}(\Pi_1(t) \cap \{1,\ldots,n\}), \\ \underline{\lambda}_1(t) & = & \liminf_{n\to\infty} n^{-1}\,\mathrm{card}(\Pi_1(t) \cap \{1,\ldots,n\}). \end{array}$$

Quand ces deux quantités coïncident, on note $\lambda_1(t)$ leur valeur commune et on dit que $\Pi_1(t)$ a une fréquence asymptotique. En fait, d'après un résultat de Kingman dans [4], pour tout $t \geq 0$ fixé, comme la partition aléatoire $\Pi(t)$ est échangeable, $\Pi_1(t)$ a p.s. une fréquence asymptotique, et $\lambda_1(t)$ est une variable aléatoire bien définie. On souhaite néanmoins étudier l'évolution du bloc $\Pi_1(t)$, t variant dans $[0,\infty[$, et non fixé dans $[0,\infty[$. Pour cela, on va introduire une nouvelle notion.

Notons $(\mathcal{F}_t)_{t\geq 0}$ la filtration naturelle associée au processus $\Pi(\cdot)$ (elle est en particulier complète). On rappelle la définition d'un subordinateur :

Définition 6 *On appelle $(\mathcal{F}_t)$-subordinateur tout processus $(\tau_t, t \geq 0)$ $(\mathcal{F}_t)$-adapté, continu à droite, à valeurs dans $[0,\infty]$, tel que $\tau_0 = 0$ p.s. et tel que pour tous $t, t' > 0$, conditionnellement à $\tau_t < \infty$, la variable aléatoire $\tau_{t+t'} - \tau_t$ est indépendante de $\mathcal{F}_t$ et a même loi que $\tau_{t'}$. La loi d'un subordinateur est caractérisée par son exposant de Laplace Φ donné par*

$$\mathbf{E}[\exp(-q\tau_t)] = \exp(-t\Phi(q)), \quad t, q > 0,$$

qui est lui-même caractérisé par la formule de Lévy-Kintchine :

$$\Phi(q) = k + dq + \int_{]0,\infty[} (1 - e^{-qx})\Lambda(dx),$$

où $k \geq 0$ est le taux de mort, $d \geq 0$ est le drift et Λ est une mesure sur $]0,\infty[$ telle que $\int (1 \wedge x)\Lambda(dx) < \infty$, appelée la mesure de Lévy de τ.

Remarquons que par convergence dominée, $\mathbf{E}[\exp(-q\tau_t)] = \mathbf{E}[\exp(-q\tau_t)\mathbf{1}_{\{\tau_t<\infty\}}]$ converge quand q tend 0 vers $\mathbf{P}(\tau_t < \infty)$, donc $\mathbf{P}(\tau_t < \infty) = e^{-kt}$: τ a une durée de vie de loi exponentielle de paramètre k.

Le résultat principal de cette section est le théorème suivant :

Théorème 3 1. *Presque sûrement, pour tout $t \geq 0$, le bloc $\Pi_1(t)$ a une fréquence asymptotique notée $\lambda_1(t)$.*

2. *Le processus $(-\log(\lambda_1(t)), t \geq 0)$ est un $(\mathcal{F}_t)$-subordinateur. Son drift est le taux d'érosion c de κ, son taux de mort est donné par*

$$k = c + \int_{\mathcal{S}^*} \Big(1 - \sum_{j=1}^{\infty} s_j\Big)\nu(ds),$$

et sa mesure de Lévy par

$$\Lambda(dx) = e^{-x} \sum_{j=1}^{\infty} \nu(-\log s_j \in dx).$$

Remarque Pour tout $j \in \mathbf{N}$, la mesure $\nu(-\log s_j \in dx)$ est la mesure image $p_{j*}\nu$ de ν par l'application $p_j : s = (s_i, i \in \mathbf{N}) \in \mathcal{S}^* \mapsto -\log s_j \in]0, \infty]$. Ainsi, pour toute fonction f mesurable positive définie sur $]0, \infty[$,

$$\int_{]0,\infty[} f d\Lambda = \sum_{j \in \mathbf{N}} \int_{]0,\infty[} e^{-x} f(x) p_{j*}\nu(dx), \quad \text{i.e.} \int_{]0,\infty[} f d\Lambda = \sum_{j \in \mathbf{N}} \int_{\mathcal{S}^*} s_j f(-\log s_j) \nu(ds).$$

La preuve de ce théorème s'effectue en trois étapes. On calcule tout d'abord l'espérance des puissances entières de $\lambda_1(t)$ à un instant t fixé. Comme on ne s'intéresse qu'à la loi du processus $\Pi(\cdot)$, on peut supposer que ce dernier est construit à partir d'un processus ponctuel de Poisson, et il est alors aisé de faire les calculs souhaités. On montre dans un deuxième temps que, en notant pour tout instant t,

$$\tau_t = -\lim_{\substack{t' \downarrow t \\ t' \in \mathbf{Q}}} \uparrow \log \lambda_1(t'),$$

le processus τ est un $(\mathcal{F}_t)$-subordinateur. Donnons l'idée de la preuve. Pour tous s et t, le bloc $\Pi_1(s+t)$ peut être vu comme le premier bloc de la restriction $\gamma \circ \Pi_1(s)$ de la partition γ au bloc $\Pi(s)$, où γ est une partition aléatoire échangeable indépendante de la tribu $\mathcal{F}_s$ et de même loi que $\Pi(t)$. Ceci montre que la fréquence asymptotique $\lambda_1(s+t)$ de $\Pi_1(s+t)$ est égale au produit $\lambda_1(s)\lambda_1'(t)$, où $\lambda_1'(t)$ est la fréquence asymptotique du premier bloc de la partition γ, distribué comme $\Pi_1(t)$ et indépendant de la tribu $\mathcal{F}_s$. On a le résultat en passant au logarithme. La dernière étape de la preuve du théorème consiste à montrer que presque sûrement, pour tout $t \geq 0$, le bloc $\Pi_1(t)$ a une fréquence asymptotique donnée par $\lambda_1(t) = \exp(-\tau_t)$.

On déduit du théorème 3 une condition nécessaire et suffisante pour que pour tout $t \geq 0$, p.s. $\lambda_1(t) > 0$. Donnons tout d'abord les définitions suivantes :

Définition 7 *On dit que le processus de fragmentations homogènes $\Pi(\cdot)$ n'a pas d'érosion si le taux d'érosion c de μ_c est nul. On dit que $\Pi(\cdot)$ est conservateur si la mesure de Lévy ν sur $\mathcal{S}^*$ associée à μ_ν vérifie :*

$$\nu\Big(\Big\{s \in \mathcal{S}^*, s_0 = 1 - \sum_{j=1}^{\infty} s_j > 0\Big\}\Big) = 0.$$

On dit aussi que la mesure ν est conservatrice.

D'après l'expression du taux de mort donnée par le théorème 3, on a : pour tout $t \geq 0$, $\lambda_1(t) > 0$ p.s. si et seulement si le processus $\Pi(\cdot)$ n'a pas d'érosion et est conservateur. Dans tout ce qui suit, on suppose que $\Pi(\cdot)$ n'a pas d'érosion et qu'il est conservateur.

On sait que presque sûrement, pour tout $t \geq 0$ et pour tout $i \in \mathbf{N}$, le bloc $\Pi_i(t)$ a une fréquence asymptotique, notée $\lambda_i(t)$. On peut montrer le résultat suivant :

Proposition 3 *Pour tout réel p strictement positif, le processus $\left(M_t^{(p)}\right)_{t\geq 0}$ défini par*

$$M_t^{(p)} = e^{\Phi(p)t} \sum_{i\in\mathbf{N}} \lambda_i^{p+1}(t)$$

est une $(\mathcal{F}_t)$-martingale.

5 Arbres de fragmentation discrète

Un *arbre* est un graphe connexe sans cycle. Il est dit enraciné si un sommet particulier de l'arbre (appelé la *racine*) est distingué, et dans ce cas les arêtes sont par convention orientées, chacune pointant dans la direction opposée à la racine. Le degré sortant d'un sommet s est le nombre de sommets adjacents à s que l'on peut rencontrer partant de s et suivant l'orientation donnée par l'arête parcourue. Une *feuille* est un sommet de degré sortant nul. Pour tout $n \in \mathbf{N}$, notons $\mathbf{T}_n$ l'ensemble des arbres enracinés avec exactement n feuilles étiquetées tels que les autres sommets (sauf la racine) ne sont pas étiquetés et tels que la racine est l'unique sommet de degré sortant égal à 1. Si $\mathbf{t} \in \mathbf{T}_n$, on note $A(\mathbf{t})$ l'ensemble de ses arêtes.

Dans tout ce qui suit, B désigne une partie finie de $\mathbf{N}$. On note $\mathbf{T}_B$ l'ensemble des collections $\mathbf{t}$ de parties de B qui contiennent aussi `RACINE` telles que :

1. $B \in \mathbf{t}$; on appelle B l'ancêtre commun,
2. $\{i\} \in \mathbf{t}$ pour tout $i \in B$,
3. pour tous $A, C \in \mathbf{t}$, $A \cap C = \emptyset$, $A \subset C$ ou $C \subset A$.

On dit que $A \in \mathbf{t}$ est un *descendant* de $C \in \mathbf{t}$ si $A \subsetneq C$. On dit alors que C est un *ancêtre* de A. Si $A \subsetneq C$ et si pour $D \in \mathbf{t}$ tel que $A \subset D \subset C$, on a $D = A$ ou $D = C$, on dit que A est un fils de C, et C est appelé le père de A. Si on munit $\mathbf{t}$ de la relation père-fils tout en liant `RACINE` à B, alors $\mathbf{t}$ est un arbre enraciné dont les feuilles sont les $\{i\}$, $i \in B$. Remarquons que l'on peut identifier $\mathbf{T}_{\{1,\ldots,n\}}$ à $\mathbf{T}_n$.

5.1 Lien avec les fragmentations homogènes

Il y a un lien naturel entre les processus à valeurs dans $\mathcal{P}_B$ et les arbres :

Définition 8 *Soit $\Pi(\cdot)$ un processus à valeurs dans $\mathcal{P}_B$ qui se raffine. Supposons que $\Pi(0)$ soit la partition triviale et qu'il existe un instant $t \geq 0$ où $\Pi(t)$ soit la partition discrète. Alors $\mathbf{t}_\Pi = \{\mathtt{RACINE}\} \cup \{A \subset B, \exists t \geq 0, A \in \Pi(t)\}$ est un élément de $\mathbf{T}_B$, appelé arbre associé à la fragmentation $\Pi(\cdot)$.*

On définit maintenant la notion fondamentale d'arbre de fragmentation discrète :

Définition 9 *Soit $\Pi(\cdot)$ un processus de fragmentations homogènes et B une partie finie de $\mathbf{N}$. On note $\Pi \circ B$ le processus $(\Pi(t) \circ B, t \geq 0)$. Alors $\Pi \circ B(t)$ est la partition discrète pour t assez grand. On définit $T_B = \mathbf{t}_{\Pi\circ B} \in \mathbf{T}_B$. On appelle un tel arbre un arbre de fragmentation discrète.*

Bien que cette représentation indique la manière dont les différents blocs se brisent, elle perd l'aspect temporel de la fragmentation. On introduit dans la suite la notion de longueur d'arête, qui permet de coder l'information sur la durée de vie des blocs.

5.2 Arbres avec des longueurs d'arête

Un arbre avec des longueurs d'arête est une paire $\mathcal{A} = (\mathbf{t}, \mathbf{l})$, où $\mathbf{t} \in \mathbf{T}_B$ et $\mathbf{l} = (l_i, i \in A(\mathbf{t}))$ est une famille de réels strictement positifs (peut-être infinis). On appelle $\mathbf{t}$ la *forme* de $\mathcal{A}$. De l'arbre $\mathcal{A}$ on déduit la distance naturelle sur l'ensemble des sommets de l'arbre : pour tous sommets v, w, on pose $d(v, w)$ la somme des longueurs des arêtes parcourues lors de l'unique chemin dans $\mathbf{t}$ allant de v à w. La *hauteur* d'un sommet est sa distance à la racine. La hauteur d'un arbre est le maximum des hauteurs des sommets de l'arbre. On note $\mathbf{A}_B$ l'ensemble des arbres avec des longueurs d'arête dont la forme est dans $\mathbf{T}_B$.

Définissons un arbre avec des longueurs d'arête dont la forme est T_B et dont les arêtes ont pour longueur la durée de vie du bloc correspondant. Remarquons qu'un sommet s est caractérisé par l'ensemble B_s des feuilles qui descendent de s. Pour tout sommet s distinct de la racine, notons D_s la durée de vie du bloc B_s :

$$D_s = \inf\{t \geq 0, \Pi(t) \circ B_s \neq \mathbf{1}_{B_s}\}.$$

Remarquons que si s est un sommet, p.s., D_s est fini si et seulement s n'est pas une feuille. On pose $D_{\texttt{RACINE}} = 0$. On note $\theta_B \in \mathbf{A}_B$ l'arbre avec des longueurs d'arête dont la forme est T_B et tel que pour tout sommet distinct de la racine, l'arête reliant s à son père $p(s)$ a pour longueur $D_s - D_{p(s)}$.

5.3 Élagage

Soit $\mathcal{A} \in \mathbf{A}_B$. On définit le processus d'*élagage* de l'arbre $\mathcal{A}$ selon un entier naturel k strictement supérieur à 1. On introduit tout d'abord la notion d'*épaisseur* d'une arête. Soit $e = (u, v)$ une arête de l'arbre. On sait que si u est différent de `RACINE`, le cardinal de u est strictement supérieur à celui de v. On appelle épaisseur de l'arête e le cardinal de v. L'arbre élagué $\mathcal{E}(\mathcal{A}, k)$ est l'arbre avec des longueur d'arête obtenu à partir de l'arbre $\mathcal{A}$ en enlevant toutes les arêtes d'épaisseur strictement inférieure à k. Ainsi, la hauteur et la *longueur totale* (i.e. la somme des longueurs des arêtes) de l'arbre $\mathcal{E}(\theta_B, k)$ sont finies. On peut s'intéresser à l'étude des comportements asymptotiques de la hauteur et de la longueur totale de l'arbre $\mathcal{E}(\theta_{\{1,\ldots,n\}}, f(n))$ lorsque n tend vers l'infini, où f est une fonction croissante (par exemple f constante, ou $f(n) = n^\alpha$ avec $\alpha \in]0, 1[$).

Références

[1] J. Bertoin, Homogeneous fragmentation processes. *Probab. Theory Related Fields*, 121 (2001), no. 3, 301-318.

[2] J. Bertoin, Self-similar fragmentations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38 (2002), no. 3, 319-340.

[3] B. Haas, G. Miermont, J. Pitman et M. Winkel, Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models, 2006.

[4] J. F. C. Kingman, The coalescent. *Stochastic Process. Appl.* 13 (1982), no. 3, 235-248.