Arbres généalogiques d'une grande population

L. de Raphelis et J. Rochet
Encadré par A. Véber

Juin 2010

Nous tenons à remercier Amandine Véber qui, en tant que directrice de mémoire, s'est toujours montrée à l'écoute et très disponible.

Table des matières

1 Introduction

1.1 Cadre

On s'intéresse à un échantillon d'individus tirés d'une population dans laquelle les individus se transmettent leurs gènes de parent à enfant. Pour simplifier, on prend des individus haploïdes (c'est-à-dire dont chaque chromosome n'apparaît qu'une fois dans le génome) qui n'ont donc qu'un parent. On cherche alors à retracer l'arbre généalogique de cet échantillon en fonction du mécanisme aléatoire suivant lequel les individus se reproduisent d'une génération à l'autre. Cet arbre est représenté ainsi : chaque génération dans le passé est elle-même représentée par une partition de l'ensemble des individus de l'échantillon, dans laquelle chaque bloc (élément de la partition) contient les individus ayant un ancêtre commun à cette génération passée (cf fig 1 pour un exemple d'arbre).

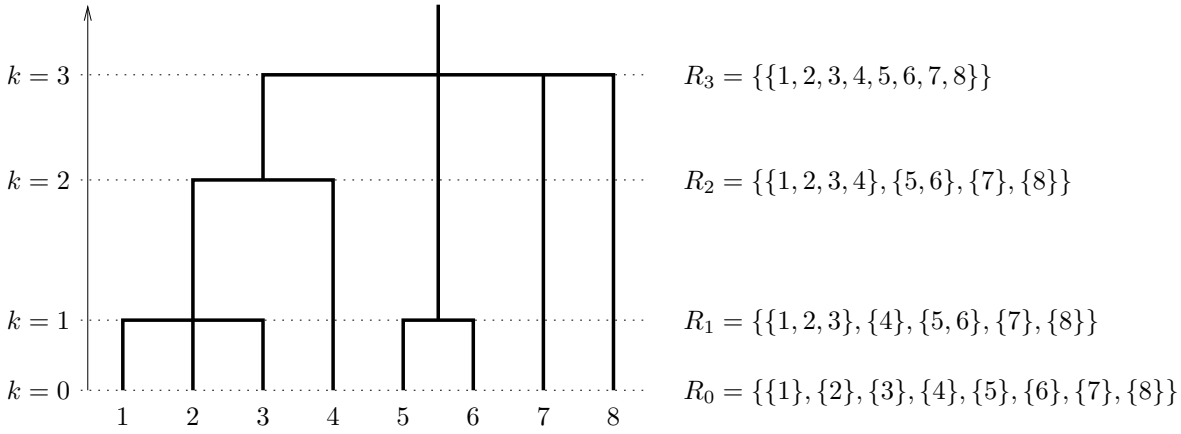

FIG. 1 – Exemple d'arbre généalogique pour n=8

Nous utilisons le modèle de reproduction suivant :

On considère une population haploïde de N individus, où N est fixé. On suit cette population sur plusieurs générations. A chaque génération, on tire aléatoirement un vecteur $\nu := (\nu_1, \dots, \nu_N)$ (vérifiant des conditions que nous allons préciser), où ν_i représente le nombre de descendants du i-ème individu. Ces vecteurs, correspondant à chaque génération, sont tirés de manière iid.
Comme on a supposé la taille de la population constante, on a la première contrainte suivante :

$$\nu_1 + \nu_2 + ... + \nu_N = N \text{ presque sûrement.}$$

On suppose de plus qu'il n'y a pas de sélection naturelle (autrement dit aucun individu n'est favorisé), ce qui se traduit par le fait que les variables ν_i sont échangeables, c'est à dire que pour toute permutation σ de $\{1, ..., N\}$

$$(\nu_1, \nu_2, ..., \nu_N) \overset{(loi)}{=} (\nu_{\sigma(1)}, \nu_{\sigma(2)}, ..., \nu_{\sigma(N)}) \ .$$

Ainsi, par exemple pour $N = 4$

$$\mathbb{P}(\nu_1 = 2, \nu_2 = 1, \nu_3 = 0, \nu_4 = 1) = \mathbb{P}(\nu_1 = 1, \nu_2 = 1, \nu_3 = 2, \nu_4 = 0).$$

De cette manière, on définit un processus aléatoire à temps discret, qui correspond au modèle de Cannings décrit dans [Can74].

On peut en donner un cas particulier : le modèle de Wright-Fisher. Dans ce modèle, chaque individu choisit son parent dans la génération précédente uniformément et indépendamment des autres. Cela correspond à prendre comme loi du vecteur ν la loi multinomiale, c'est à dire pour des entiers $n_1, \ldots, n_N$ tels que $n_1 + ... + n_N = N$:

$$\mathbb{P}(\nu_1 = n_1, \ldots, \nu_N = n_N) = \frac{N!}{n_1! \ldots n_N!} \prod_{i=1}^{N} \left(\frac{1}{N}\right)^{n_i} = \frac{N!}{n_1! \ldots n_N!} \frac{1}{N^N}.$$

Revenons au cas général ; on fixe $n \leq N$ qui désigne la taille d'un échantillon d'individus pris aléatoirement parmi les N individus de la population initiale.

Notre objectif est d'étudier le comportement de l'arbre généalogique de l'échantillon de taille n (fixée) lorsque N devient de plus en plus grand. Pour retracer cet arbre, nous serons amenés à remonter dans le passé et donc à considérer des générations suivant un temps allant dans le sens inverse du sens habituel (l'instant initial sera le temps 0).

Représentation de l'arbre : On note $(R_k^N)_{k\in\mathbb{N}}$ la chaîne de Markov ayant pour espace d'états $E_n = \{partitions\ de\ \ \{1, ..., n\}\}$, partant de $R_0^N = \{\{1\}, ..., \{n\}\}$ (état initial) et où R_k^N représente les relations généalogiques entre les individus de l'échantillon k générations dans le passé : i et j appartiennent au même bloc de R_k^N si et seulement s'ils ont un ancêtre commun à la $k^{ème}$ génération dans le passé. Par exemple (pour $n = 6$) si

$$R_k^N = \{\{1, 3, 6\}; \{2\}; \{4, 5\}\} \ ,$$

cela signifie que, dans les k générations antérieures, les individus 1, 3 et 6 atteignent leur ancêtre commun, de même pour les individus 4 et 5, tandis que l'individu 2 n'en partage aucun avec les autres (voir fig 1).

Introduisons la relation de comparaison réflexive $\subseteq$ sur E_n définie par $\forall \xi, \eta \in E_n,\ \xi \subseteq \eta$ si et seulement si la partition η peut s'obtenir en fusionnant certains blocs de ξ. Par exemple

$$\{\{1, 3\}, \{2\}, \{4, 6\}, \{5\}, \{7\}\} \subseteq \{\{1, 3, 4, 6\}, \{2, 7\}, \{5\}\} \ .$$

On peut remarquer que $\forall \xi \in E_n$ on a

$$\{\{1\}, \{2\}, ..., \{n\}\} \subseteq \xi \subseteq \{\{1, ..., n\}\} \ .$$

On note $p_{\xi,\eta} = P(R_k^N = \eta | R_{k-1}^N = \xi)$ la probabilité de transition de l'état ξ à η. On a évidement que $p_{\xi,\eta} = 0$ si $\xi \subsetneq \eta$, car lorsqu'on remonte une génération, les blocs ne peuvent que fusionner. On note aussi c_N la probabilité que deux individus pris au hasard aient le même ancêtre à la génération précédente.

Remarque : La probabilité c_N est appelée *probabilité de coalescence*. Elle joue un rôle central dans l'étude des coalescents. En effet, on comprend facilement que la probabilité que trois individus ou plus aient le même parent est plus petite que c_N. On pourra donc utiliser c_N comme point de comparaison pour les $p_{\xi,\eta}$.

Introduisons maintenant la définition suivante :

Définition 1. *Soient $\xi \in E_n$ et $\eta \in E_n$, où $\xi \subseteq \eta$. On dit que ξ est une k-fusion de η, ce que l'on notera $\xi \prec_k \eta$, si et seulement si η s'obtient à partir de ξ en fusionnant k des blocs de ce dernier en un seul, et en laissant les autres inchangés.*

Par exemple, $\{\{1,3\},\{2\},\{4,6\},\{5\},\{7\}\} \prec_3 \{\{1,3,4,5,6\},\{2\},\{7\}\}$.

Remarque : On verra que cette notion de k-fusion est suffisante pour décrire les événements entre deux générations. En effet sous certaines conditions, lorsque N est grand, la probabilité que plusieurs groupes fusionnent simultanément en deux groupes distincts ou plus est négligeable devant c_N et donc ce type de transition n'apparaît plus à la limite.

1.2 Théorème de convergence principal

L'objectif du théorème suivant est précisément d'étudier le comportement de la chaîne de Markov $(R_k^N)_{k\in\mathbb{N}}$, lorsque N tend vers $+\infty$ et que la taille n de l'échantillon est fixée. Ce théorème est l'objet principal de l'article de Möhle et Sagitov [MS01].

Théorème 1. *Si l'on suppose les deux conditions suivantes satisfaites :*

$$(C1) \quad \phi_k := \lim_{N\to\infty} \frac{N^{1-k}}{c_N}\mathbb{E}[(\nu_1)_k] \text{ existe pour tout } k \geq 2.$$
$$(C2) \quad \lim_{N\to\infty} \frac{1}{N^2 c_N}\mathbb{E}[(\nu_1)_2(\nu_2)_2] = 0$$

Alors pour tout échantillon de taille $n \in \mathbb{N}$, la limite suivante existe (en notant $P_N = (p_{\xi,\eta})_{\xi,\eta\in E_n}$ et I_d la matrice $(\delta_{\xi,\eta})_{\xi,\eta\in E_n}$),

$$Q = (q_{\xi,\eta})_{\xi,\eta\in E_n} := \lim_{N\to\infty} \frac{P_N - I_d}{c_N}$$

telle que pour tout $\xi \in E_n$ ayant b blocs et tout $\eta \in E_n$

$$q_{\xi,\eta} = \begin{cases} - \int_{[0,1]} \frac{1-(1-x)^{b-1}(1-x+bx)}{x^2}\Lambda(dx) & si\ \xi = \eta, \\ \int_{[0,1]} x^{k-2}(1-x)^{b-k}\Lambda(dx) & si\ \xi \prec_k \eta, \\ 0 & sinon. \end{cases}$$

Ici, Λ est une mesure de probabilité sur $[0,1]$ uniquement déterminée par ses moments

$$\int_{[0,1]} x^k \Lambda(dx) = \phi_{k+2},\ k \in \mathbb{N}.$$

Remarque (*) : Ces conditions sont vérifiées entre autres par le modèle de Wright-Fisher. En effet, on a :

$$\begin{aligned} c_N &= \frac{1}{N}, \\ \mathbb{E}[(\nu_1)_k] &\to 1 \text{ quand } N \to \infty, \\ \mathbb{E}[(\nu_1)_2(\nu_2)_2] &\to 1 \text{ quand } N \to \infty. \end{aligned}$$

Ce modèle correspond au cas simple où $\forall k \geq 3,\ \phi_k = 0$.
(Une démonstration de cette remarque est proposée en annexe).

1.3 Le coalescent

Nous allons nous intéresser dans ce qui suit à un certain objet mathématique, appelé coalescent, qui permet également de décrire l'évolution des relations généalogiques entre des individus à travers le temps. Il a été initialement introduit par J.F.C Kingman (Cf [Kin82]). Il s'agit un processus de Markov à temps continu, commençons donc par définir ce type d'objets.

Définition 2. *Un processus $(X(t))_{t\in\mathbb{R}_+}$ à valeurs dans un ensemble dénombrable S a la propriété de Markov si l'égalité*

$$\mathbb{P}(X(s+t)=y|X(s)=x, X(s_n)=x_n,\dots,X(s_0)=x_0)=\mathbb{P}(X(s+t)=y|X(s)=x)$$

est satisfaite pour tous $n\in\mathbb{N}$, $0\leq s_0\leq s_1\leq\ldots\leq s_n\leq s$, $t\geq 0$, $x,y\in S$ et $x_0,\dots,x_n\in S$.
De façon équivalente, pour $s\leq t$ on a l'égalité des distributions

$$\mathbb{P}(X(s+t)=x|\mathcal{F}_s)=\mathbb{P}(X(s+t)=x|X(s))\ \forall x\in S\ ,$$

où $\mathcal{F}_s$ est la filtration engendrée par $\{X_u,\ s\leq t\}$.
Un processus vérifiant la propriété de Markov est dit processus de Markov. Il est dit homogène si en outre la propriété

$$\mathbb{P}(X(s+t)=x_n|X(s)=x_{n-1})=\mathbb{P}(X(t)=x_n|X(0)=x_{n-1})$$

est vérifiée pour tous s, t, x_{n-1} et x_n.

Les processus de Markov forment une famille très générale de processus ; en ce qui nous concerne, l'évolution d'une population en suivant les générations se fait, évidemment, suivant un temps continu, mais les événements arrivent de façon discrète. Ceci correspond à ce que l'on appelle les processus de sauts, définis de la manière suivante :

Définition 3. *Soient*
— $P=(p(x,y))_{x,y\in S}$ la matrice de transition d'une chaîne de Markov $((M_n))_{n\in\mathbb{N}}$ telle que p(x,x)=0 pour tout $x\in S$;
— $(q_x)_{x\in S}$ une suite de réels strictement positifs ;
— $(t_n)_{n\in\mathbb{N}}$ une suite croissante de variables aléatoires à valeurs réelles définie par récurrence par :

$$t_0=0,\quad t_{n+1}=t_n+E^n_{M_n},\quad n\geq 0$$

où, pour tout $x\in S$, $(E^n_x)_{n\geq 0}$ est une suite de variables exponentielles de paramètre q_x.
On suppose que presque sûrement, la suite (t_n) tend vers l'infini.
Alors le processus $(X(t))_{t\geq 0}$, appelé processus de sauts, est défini par

$$X(t)=M_n\qquad si\ t_n\leq t<t_{n+1}.$$

Il est caractérisé par sa matrice de sauts $Q=(q(x,y),(x,y)\in S^2)$, donnée par

$$q(x,y)=q_xp(x,y)\ si\ x\ \neq y\ et\ q(x,x)=-q_x.$$

On peut voir un tel processus ainsi : on attend un temps de loi exponentielle (le paramètre dépend de l'état x dans lequel on est), puis on change d'état suivant la loi $p(x,.)$ donnée par la matrice P.
Un exemple de processus de sauts est la marche aléatoire simple à temps continu sur $\mathbb{Z}$: on attend un temps de loi exponentielle (de paramètre fixé) après chaque saut, puis avec équiprobabilité on va à gauche ou à droite.
On pourra vérifier qu'un processus de sauts a bien la propriété de Markov.

Enfin, on appelle coalescent échangeable un processus de sauts à valeurs dans les partitions de $\mathbb{N}$ dont l'évolution ne peut se faire que par fusion de blocs ; de plus, celle-ci ne doit pas dépendre des blocs (d'où le terme «échangeable», qui s'applique aux blocs).
Définissons maintenant un cas particulier de coalescent, le Λ-coalescent (ou coalescent à collisions multiples) :

Définition 4. (Λ-*coalescent*)
Soit Λ une mesure finie sur $[0;1]$. On appelle coalescent à collisions multiples de mesure Λ, ou Λ-coalescent, le processus markovien à valeurs dans les partitions de $\mathbb{N}$ tel que pour tout $n \in \mathbb{N}$, sa restriction à E_n est un processus markovien dont les taux de transition q_Λ sont donnés par : si $\xi, \eta \in E_n$, $b := |\xi|$ et $\xi \prec_k \eta$ pour un $k \in \{2, \dots, b\}$, alors

$$q_\Lambda(\xi, \eta) = \int_{[0;1]} x^k (1-x)^{b-k} \frac{\Lambda(dx)}{x^2},$$

et les autres taux de transition sont nuls.

Un exemple simple de Λ-coalescent est le coalescent de Kingman, pour lequel on prend pour mesure $\Lambda = \delta_0$ (la mesure de Dirac en 0), ce qui donne les taux de transition suivants :

$$q_K(\xi, \eta) = \begin{cases} 1 & si\ \xi \prec_2 \eta, \\ 0 & sinon. \end{cases}$$

1.4 Convergence des coalescents associés

Ce second théorème est une conséquence naturelle du théorème 1. Il a pour but de montrer qu'en procédant à un changement d'échelle de temps (c'est à dire en étudiant le processus de Markov à temps continu $(R^N_{[t/c_N]})_{t\geq 0}$, ce qui revient en fait à compter les générations par paquets de c_N^{-1} unités de temps initiales) le processus $(R^N_{[t/c_N]})_{t\geq 0})$ converge, dans un sens à définir, vers un Λ-coalescent. Par exemple pour le modèle de Wright-Fisher, le processus tend vers le coalescent de Kingman. Voici l'énoncé du second théorème :

Théorème 2. *On suppose que les conditions $(C1)$ et $(C2)$ du théorème 1 sont satisfaites et qu'en outre $c := \lim_{N\to\infty} c_N$ existe. Alors*

— *Si $c > 0$, quand $N \to \infty$ le processus $(R^N_k)_{k\in\mathbb{N}}$ converge vers un processus de Markov à temps discret $(\mathcal{R}_k)_{k\in\mathbb{N}}$, ayant pour état initial $\mathcal{R}_0 = \{\{1\}, ..., \{n\}\}$ et pour matrice de transition $I_d + cQ$ où Q est donnée dans le théorème 1.*

— *Si $c = 0$, quand $N \to \infty$ le processus à temps continu $(R^N_{[t/c_N]})_{t\geq 0}$ converge vers un Λ-coalescent $(\mathcal{R}_t)_{t\geq 0}$, ayant pour état initial $\mathcal{R}_0 = \{\{1\}, ..., \{n\}\}$ et de matrice de sauts Q, où Q est donnée dans le théorème 1.*

Ici la convergence est la convergence en loi des marginales fini-dimensionnelles.

1.5 Applications à la biologie

Les principales applications de ces deux théorèmes concernent le domaine de la génétique des populations. En effet, celle-ci a pour but de comprendre les mécanismes qui créent et maintiennent la variation génétique au sein des espèces. Ces mécanismes sont par exemple la mutation, la sélection naturelle, l'aléa des naissances et des morts. La génétique des populations étudie entre autres la propagation ou la disparition d'un certain gène dans une population, ce qui peut permettre de retracer l'histoire de celle-ci, ses flux migratoires, ou au contraire prédire les variations de fréquence d'un gène dans la population.

Cependant, la génétique met en jeu le traitement de données de taille gigantesque ; il est donc nécessaire de prendre des modèles relativement simples permettant de comprendre, d'interpréter et

de prédire ces données. Le coalescent constitue en cela un modèle simple qui est déjà au coeur de beaucoup de travaux en génétique des populations : dans le cadre d'une expérience le mettant en jeu, il n'est ainsi pas nécessaire de suivre toute une population (taille de l'ordre N), mais seulement un échantillon (de l'ordre de $n << N$). On compare ensuite cet échantillon avec le coalescent : dans le cas d'une évolution normale, on observera une distribution des allèles qui correspond avec celle du coalescent sur lequel on a greffé des mutations à un certain taux. Dans le cas contraire, on peut déduire l'influence probable de certains phénomènes tels qu'une migration (si la distribution des gènes est clairement scindée en 2 groupes par exemple), un bouleversement ayant occasionné une réduction drastique de la taille de la population (on observera par exemple des ancêtres communs beaucoup plus récents que prévu si la population s'est vue réduite à un très petit nombre d'individus, leurs allèles étant communs à beaucoup de leur descendants très rapidement), ou encore la non-neutralité d'un gène donné (s'il est beaucoup plus présent dans la population que ne l'aurait prédit le modèle du coalescent, alors on peut supposer que ce gène a été favorisé par la sélection naturelle).
Par ailleurs, on pourra remarquer que l'on traite dans cet exposé d'individus haploïdes ; cependant le cas diploïde (par exemple les populations humaines) pourra se traiter sous certaines conditions en considérant une population de taille 2N (voir [Wak08]).

2 Preuve du théorème de convergence principal

Commençons par déterminer l'expression des $p_{\xi,\eta}$ et celle de c_N. Si $\xi \subseteq \eta$, notons $B_1, ..., B_a$ les différents blocs de η où a désigne son nombre de blocs, et $B_{\alpha,\beta},\ \alpha \in \{1, ..., a\},\ \beta \in \{1, ..., b_\alpha\}$ ceux de ξ de sorte que $B_\alpha = \bigcup_{\beta=1}^{b_\alpha} B_{\alpha,\beta}$ pour tout α.

Lemme 1. *La probabilité de transition $p_{\xi,\eta}$ est donnée par*

$$p_{\xi,\eta} = \frac{1}{(N)_b} \sum_{\substack{i_1,i_2,...,i_a=1 \\ tous\ distincts}}^{N} \mathbb{E}[(\nu_{i_1})_{b_1}...(\nu_{i_a})_{b_a})] = \frac{(N)_a}{(N)_b}\mathbb{E}[(\nu_1)_{b_1}...(\nu_a)_{b_a}],$$

où $b := b_1 + b_2 + ... + b_a$ désigne le nombre de blocs de ξ et où on utilise la notation $(x)_p = x(x-1)...(x-p+1)$.

Preuve : Commençons par calculer c_N :

$$c_N = \sum_{i=1}^{N} \mathbb{P}[\ parent\ commun\ =\ i\ ] = \sum_{i=1}^{N} \mathbb{E}\left[\frac{\nu_i}{N}\frac{(\nu_i - 1)}{(N-1)}\right] = \frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}[(\nu_i)_2] = \frac{1}{(N-1)}\mathbb{E}[(\nu_i)_2],$$

où la dernière égalité provient de l'échangeabilité des variables ν_i. En effet, pour que le parent commun soit bien i, il faut avoir choisi le premier individu parmi les ν_i descendants de i et le second doit être choisi dans les $N-1$ individus restants, parmi les $\nu_i - 1$ autres descendants. Ainsi, on peut généraliser le raisonnement précédent avec la probabilité que a groupes aient tous un ancêtre différent,

$$p_{\xi,\eta} = \sum_{\substack{i_1,i_2,...,i_a=1 \\ tous\ distincts}}^{N} \mathbb{E}\left[\frac{(\nu_{i_1})_{b_1}}{(N)_{b_1}}\frac{(\nu_{i_2})_{b_2}}{(N-b_1)_{b_2}}...\frac{(\nu_{i_a})_{b_a}}{(N-b_1-b_2-...-b_{a-1})_{b_a}}\right] = \frac{(N)_a}{(N)_b}\mathbb{E}[(\nu_1)_{b_1}...(\nu_a)_{b_a}]\ ,$$

où la dernière égalité provient toujours de l'échangeabilité et en utilisant le fait que

$$Card(\{(i_1, ..., i_a) \in \{1, ..., N\}^a\ tel\ que\ \forall j, k \in \{1, ..., N\},\ i_j \neq i_k\}) = (N)_a$$

et

$$(N)_{b_1}(N-b_1)_{b_2}...(N-b_1-b_2-...-b_{a-1})_{b_a} = (N)_b$$

.

□

Remarque : La condition $(C2)$ est satisfaite par exemple si $\lim_{N\to\infty} c_N = 0$ et si $Cov((\nu_1)_2, (\nu_2)_2) \leq 0$. En effet, on a alors

$$\mathbb{E}[(\nu_1)_2(\nu_2)_2] \leq (\mathbb{E}[(\nu_1)_2])^2 \sim N^2 c_N^2,$$

donc

$$\frac{1}{N^2 c_N}\mathbb{E}[(\nu_1)_2(\nu_2)_2] \to 0.$$

La démonstration de ce théorème 1 se fera en trois parties.

2.1 Cas où ξ n'est pas une k-fusion de η

Pour traiter ce cas, nous avons besoin du lemme suivant :

Lemme 2. *La condition $(C2)$ est équivalente à*

$$(C2)' \quad \lim_{N\to\infty} \frac{N^{a-b}}{c_N}\mathbb{E}[(\nu_1)_{b_1}...(\nu_a)_{b_a}] = 0$$

pour tout $a \geq 2$, $b_1, b_2 \geq 2$, et $b_3, ..., b_a \geq 1$, où $b := b_1 + ... + b_a$.

Preuve : $(C2)' \Rightarrow (C2)$:
Il suffit de prendre $a = b_1 = b_2 = 2$, on a exactement la condition $(C2)$.
$(C2) \Rightarrow (C2)'$:
On a

$$\begin{aligned}
S &:= \sum_{\substack{i_1,i_2,...,i_a=1 \\ tous\ distincts}}^{N} (\nu_{i_1})_{b_1}...(\nu_{i_a})_{b_a} \\
&\leq \sum_{\substack{i_1,i_2=1 \\ i_1\neq i_2}}^{N} (\nu_{i_1})_2 \nu_{i_1}^{b_1-2} (\nu_{i_2})_2 \nu_{i_2}^{b_2-2} \sum_{\substack{i_3,...,i_a=1 \\ tous\ distincts}}^{N} (\nu_{i_3})_{b_3}...(\nu_{i_a})_{b_a} \\
&\leq \sum_{\substack{i,j=1 \\ i\neq j}}^{N} (\nu_i)_2 N^{b_1-2} (\nu_j)_2 N^{b_2-2} (\nu_3 + ... + \nu_N)^{b_3+...+b_a} \\
&\leq N^{b-4} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} (\nu_i)_2 (\nu_j)_2.
\end{aligned}$$

Par ailleurs $\nu_i \leq \nu_1 + ... + \nu_N = N$ et

$$\sum_{\substack{i_3,...,i_a=1 \\ tous\ distincts}}^{N} (\nu_{i_3})_{b_3}...(\nu_{i_a})_{b_a} \leq \sum_{\substack{i_3,...,i_a=1 \\ tous\ distincts}}^{N} (\nu_{i_3})^{b_3}...(\nu_{i_a})^{b_a} \leq (\nu_3 + ... + \nu_N)^{b_3+...+b_N} \leq N^{b-b_1-b_2}.$$

On a donc

$$\mathbb{E}[S] \leq N^{b-4} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} \mathbb{E}[(\nu_i)_2(\nu_j)_2] \leq N^{b-2}\mathbb{E}[(\nu_1)_2(\nu_2)_2].$$

Ainsi

$$\begin{aligned}\frac{N^{a-b}}{c_N}\mathbb{E}[(\nu_1)_{b_1}...(\nu_a)_{b_a}] &\sim \frac{(N)_a}{(N)_b c_N}\mathbb{E}[(\nu_1)_{b_1}...(\nu_a)_{b_a}]\\ &= \frac{1}{(N)_b c_N}\mathbb{E}[S]\\ &\leq \frac{N^{b-2}}{(N)_b c_N}\mathbb{E}[(\nu_1)_2(\nu_2)_2]\\ &\sim \frac{1}{N^2 c_N}\mathbb{E}[(\nu_1)_2(\nu_2)_2]\\ &\to 0,\end{aligned}$$

ce qui nous donne bien la condition $(C2)'$. □

Remarque : On comprend bien que la condition $(C2)$ du théorème signifie que la probabilité que deux 2-fusions aient lieu simultanément est négligeable devant c_N lorsque N est grand, alors que la condition $(C2)'$ stipule que la probabilité qu'une b_1-fusion, une b_2-fusion, ... et une b_a-fusion aient lieu simultanément est négligeable devant c_N quand N est grand. Le lemme 2 devient alors intuitif et il est facile de voir que finalement la condition $(C2)$ est équivalente à la proposition suivante :

$$q_{\xi,\eta} := \lim_{N\to\infty}\frac{p_{\xi,\eta}}{c_N} = 0$$

pour tout $n \in \mathbb{N}$ et pour tout $\xi, \eta \in E_n$ avec $\xi \subseteq \eta$ mais ξ n'est pas une k-fusion de η.
Ceci conclut le cas où ξ n'est pas une k-fusion de η.

2.2 Existence de la mesure de probabilité Λ

Pour cela, démontrons le lemme suivant :

Lemme 3. *Si l'on suppose la condition $(C1)$ satisfaite, alors il existe une mesure de probabilité Λ sur $[0,1]$ telle que :*

$$\phi_k = \lim_{N\to\infty}\frac{N^{1-k}}{c_N}\mathbb{E}[(v_1)_k] = \int_{[0,1]} x^{k-2}\Lambda(dx)$$

pour tout $k \geq 2$, et telle que Λ soit entièrement caractérisée par ses moments.

Preuve : Soit Y_N la variable aléatoire à valeurs entières de loi :

$$\mathbb{P}(Y_N = i) = \frac{(i)_2}{\mathbb{E}[(\nu_1)_2]}\mathbb{P}(\nu_1 = i) = \frac{i(i-1)}{(N-1)c_N}\mathbb{P}(\nu_1 = i)$$

où $i \in \{0,\dots,N\}$. Le k^{eme} moment de $X_N = \frac{Y_N}{N}$ est :

$$\begin{aligned}\mathbb{E}[X_N^k] &= \sum_{i=0}^{N}\Big(\frac{i}{N}\Big)^k\mathbb{P}(Y_N = i)\\ &= \sum_{i=0}^{N}\Big(\frac{i}{N}\Big)^k\frac{i(i-1)}{(N-1)c_N}\mathbb{P}(\nu_1 = i)\\ &= \frac{N^{-k}}{(N-1)c_N}\mathbb{E}[\nu_1^{k+2} - \nu_1^{k+1}]\end{aligned}$$

En utilisant le fait que $t^k = \sum_{l=1}^k S_{k,l}(t)_l$ pour tout $t \in \mathbb{R}$ et tout $k \geq 1$, où $S_{k,l}$ sont les nombres de Stirling de seconde espèce, on obtient donc :

$$\begin{aligned}
\mathbb{E}[X_N^k] &= \frac{N^{-k}}{(N-1)c_N}\mathbb{E}\Big[\sum_{l=1}^{k+2} S_{k+2,l}(\nu_1)_l - \sum_{l=1}^{k+1} S_{k+1,l}(\nu_1)_l\Big] \\
&= \frac{N^{-k}}{(N-1)c_N}\mathbb{E}\Big[(\nu_1)_{k+2} + \sum_{l=2}^{k+1}(S_{k+2,l} - S_{k+1,l})(\nu_1)_l\Big] \\
&= \frac{N^{-k}}{(N-1)c_N}\mathbb{E}[(\nu_1)_{k+2}] + \sum_{l=2}^{k+1}(S_{k+2,l} - S_{k+1,l})\frac{N^{-k}}{(N-1)c_N}\mathbb{E}[(\nu_1)_l]
\end{aligned}$$

D'après la condition (C1), pour N grand chaque terme de la somme est équivalent à $(S_{k+2,l} - S_{k+1,l})\frac{\phi_l}{N^{k+2-l}}$ donc tend vers zéro, et le membre de gauche converge vers ϕ_{k+2} ; $\mathbb{E}[X_N^k]$ converge donc vers ϕ_{k+2}. On note que $0 \leq X_N \leq 1$ presque sûrement ; ainsi comme $[0;1]$ est compact, en utilisant le théorème de Stone-Weierstrass, la convergence des moments de X_N implique la convergence faible de X_N vers une limite X. Ainsi la distribution $\Lambda := P_X$ de X est déterminée de manière unique par ses moments $\mathbb{E}[X^k] = \lim_{N\to\infty}\mathbb{E}[X_N^k] = \phi_{k+2}$. C'est-à-dire :

$$\phi_{k+2} = \mathbb{E}[X^k] = \int_{[0;1]} x^k \Lambda(dx)$$

Vérifions pour finir que Λ est bien de masse totale 1. Puisque $\mathbb{E}[(\nu_1)_2] = (N-1)c_N$, on en déduit que $\phi_2 = \lim_{N\to\infty}\frac{1}{Nc_N}\mathbb{E}[(\nu_1)_2] = 1$ ce qui assure que Λ est bien une mesure de probabilité.

2.3 Cas où $\xi \prec_k \eta$

Commençons par démontrer le lemme suivant :

Lemme 4. *Si l'on suppose que les conditions* $(C1)$ *et* $(C2)$ *du théorème sont satisfaites, alors*

$$\lim_{N\to\infty}\frac{N^{1-k}}{c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a] = \int_{[0,1]} x^{k-2}(1-x)^{a-1}\Lambda(dx)$$

pour tout $k \geq 2$ *et tout* $a \geq 1$.

Preuve : On procède par récurrence sur a. Pour $a = 1$, il s'agit du lemme 3. Supposons maintenant la propriété vraie pour un certain $a \geq 1$ et pour tout $k \geq 2$. Tout d'abord notons que

$$\sum_{i=a+1}^{N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a\nu_i] = (N-a)\mathbb{E}[(\nu_1)_k\nu_2...\nu_a\nu_{a+1}] \sim N\mathbb{E}[(\nu_1)_k\nu_2...\nu_a\nu_{a+1}],$$

ceci grâce à l'échangeabilité des variables ν_i. On a d'autre part :

$$\begin{aligned}
\sum_{i=a+1}^{N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a\nu_i] &= \mathbb{E}[(\nu_1)_k\nu_2...\nu_a(\nu_{a+1} + ... + \nu_N)] \\
&= \mathbb{E}[(\nu_1)_k\nu_2...\nu_a(N - \nu_1 - ... - \nu_a)] \\
&= \mathbb{E}[(\nu_1)_k\nu_2...\nu_a((N-k-a+1) - (\nu_1 - k) - \sum_{i=2}^{a}(\nu_i - 1))] \\
&= (N-k-a+1)\mathbb{E}[(\nu_1)_k\nu_2...\nu_a] - \mathbb{E}[(\nu_1)_{k+1}\nu_2...\nu_a] \\
&\quad -(a-1)\mathbb{E}[(\nu_1)_k(\nu_2)_2\nu_3...\nu_a].
\end{aligned}$$

La dernière ligne utilise encore l'échangeabilité des ν_i.
En multipliant par $\frac{N^{-k}}{c_N}$ et en prenant la limite $N \to \infty$, on obtient alors :

$$\begin{aligned}\lim_{N\to\infty} \frac{N^{1-k}}{c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a\nu_{a+1}] &= \lim_{N\to\infty} \frac{N^{1-k}}{c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a] - \lim_{N\to\infty} \frac{N^{-k}}{c_N}\mathbb{E}[(\nu_1)_{k+1}\nu_2...\nu_a] \\ &\quad -(a-1)\lim_{N\to\infty} \frac{N^{-k}}{c_N}\mathbb{E}[(\nu_1)_k(\nu_2)_2\nu_3...\nu_a],\end{aligned}$$

D'après le lemme 2, la dernière limite est nulle. Ainsi,

$$\begin{aligned}\lim_{N\to\infty} \frac{N^{1-k}}{c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a\nu_{a+1}] &= \lim_{N\to\infty} \frac{N^{1-k}}{c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a] - \lim_{N\to\infty} \frac{N^{-k}}{c_N}\mathbb{E}[(\nu_1)_{k+1}\nu_2...\nu_a] \\ &= \int_{[0,1]} x^{k-2}(1-x)^{a-1}\Lambda(dx) - \int_{[0,1]} x^{k-1}(1-x)^{a-1}\Lambda(dx) \\ &= \int_{[0,1]} x^{k-2}(1-x)^{a}\Lambda(dx),\end{aligned}$$

ce qui valide la propriété au rang $a+1$, et termine la récurrence. □

Corollaire 1. *Soient ξ et $\eta \in E_n$. Si l'on suppose les conditions $(C1)$ et $(C2)$ vérifiées et si $\xi \prec_k \eta$ pour un certain $k \geq 2$, alors*

$$q_{\xi,\eta} := \lim_{N\to\infty} \frac{p_{\xi,\eta}}{c_N} = \int_{[0,1]} x^{k-2}(1-x)^{a-1}\Lambda(dx),$$

où a est le nombre de blocs de η.

Preuve : En notant a (respectivement b) le nombre de blocs de η (respectivement ξ), on a

$$\begin{aligned}q_{\xi,\eta} &= \lim_{N\to\infty} \frac{p_{\xi,\eta}}{c_N} \\ &= \lim_{N\to\infty} \frac{(N)_a}{(N)_b c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a] \\ &= \lim_{N\to\infty} \frac{N^a}{N^b c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a] \\ &= \lim_{N\to\infty} \frac{N^{1-k}}{c_N}\mathbb{E}[(\nu_1)_k\nu_2...\nu_a]\end{aligned}$$

puisque $a = b - k + 1$. On conclut grâce au lemme 4. □

Il ne reste plus que le cas $\xi = \eta$. Soit donc $\xi \in E_n$ et notons b son nombre de blocs. On a

$$\sum_{\eta\in E_n} p_{\xi,\eta} = 1$$

d'où il découle que

$$
\begin{aligned}
q_{\xi,\xi} &:= \lim_{N\to\infty} \frac{p_{\xi,\xi}-1}{c_N} \\
&= \lim_{N\to\infty} -\sum_{\substack{\eta\in E_n \\ \eta\neq\xi}} \frac{p_{\xi,\eta}}{c_N} \\
&= -\sum_{\substack{\eta\in E_n \\ \eta\neq\xi}} q_{\xi,\eta} = -\sum_{k=2}^{b} \sum_{\substack{\eta\in E_n \\ \xi\prec_k\eta}} q_{\xi,\eta} \\
&= -\sum_{k=2}^{N} \binom{b}{k} \int_{[0,1]} x^{k-2}(1-x)^{b-k}\Lambda(dx) \\
&= -\int_{[0,1]} x^{-2} \sum_{k=2}^{b} \binom{b}{k} x^k(1-x)^{b-k}\Lambda(dx) \\
&= -\int_{[0,1]} x^{-2}(1-(1-x)^b - bx(1-x)^{b-1})\Lambda(dx) \\
&= -\int_{[0,1]} \frac{1-(1-x)^{b-1}(1-x+bx)}{x^2}\Lambda(dx).
\end{aligned}
$$

□

Ceci termine la démonstration du théorème 1.

Remarque : On peut étendre la notion de coalescent échangeable en autorisant les fusions simultanées de plusieurs groupes de blocs. On obtient alors des Ξ-coalescents, qui généralisent la notion de Λ-coalescent. Dans ce cas, le théorème 1 peut être adapté en modifiant la condition $(C2)$ en

$$
(\hat{C}2) \qquad \lim_{N\to\infty} \frac{N^{a-b}}{c_N}\mathbb{E}[(\nu_1)_{b_1}...(\nu_a)_{b_a}] \text{ existe pour } a\geq 2, \text{ et } b\geq 2 \text{ avec } b := b_1+...+b_a.
$$

Dans ce cas, les fusions simultanées ne sont plus (forcément) négligeables devant c_N. Le résultat auquel on arrive alors est la convergence de $\frac{P_N-I_d}{c_N}$ vers une matrice Q caractérisant les Ξ-coalescents. On pourra se référer à [Sch00] et à [Sag03] pour une étude approfondie des coalescents dits à collisions multiples et simultanées (ou Ξ-coalescent).

3 Preuve du théorème de convergence des coalescents

Rappelons l'énoncé du théorème 2 :

Théorème 2. *On suppose que les conditions $(C1)$ et $(C2)$ du théorème 1 sont satisfaites et qu'en outre $c := \lim_{N\to\infty} c_N$ existe. Alors*

— *Si $c>0$, quand $N\to\infty$ le processus $(R_k^N)_{k\in\mathbb{N}}$ converge vers un processus de Markov à temps discret $(\mathcal{R}_k)_{k\in\mathbb{N}}$, ayant pour état initial $\mathcal{R}_0=\{\{1\},...,\{n\}\}$ et pour matrice de transition I_d+cQ où Q est donnée dans le théorème 1.*

— *Si $c=0$, quand $N\to\infty$ le processus à temps continu $(R^N_{[t/c_N]})_{t\geq 0}$ converge vers un Λ-coalescent $(\mathcal{R}_t)_{t\geq 0}$, ayant pour état initial $\mathcal{R}_0=\{\{1\},...,\{n\}\}$ et de matrice de sauts Q, où Q est donnée dans le théorème 1.*

Ici la convergence est la convergence en loi des marginales fini-dimensionnelles.

Preuve : D'après le théorème 1, on a bien $Q := \lim_{N\to\infty} \frac{P_N - I_d}{c_N}$ de la forme voulue.
Commençons par le cas $c > 0$. Puisque $P_N = I_d + c_N Q + o(c_N)$, on a immédiatement que $\lim_{N\to\infty} P_N = I_d + cQ$, et donc toutes les itérées P_N^k de P_N convergent vers $(I_d + cQ)^k$, ce qui suffit pour conclure.
Supposons désormais que $c = 0$. Du fait que que $c_N \to 0$, on a pour tout $t \geq 0$:

$$||P_N^{[t/c_N]} - (I_d + c_N Q)^{[t/c_N]}|| \leq [t/c_N]||P_N - (I_d + c_N Q)||,$$

où $||M||$ est la norme infinie de la matrice M. Or $P_N - (I_d + c_N Q) = o(c_N)$ (par le théorème 1), donc

$$\lim_{N\to\infty} ||P_N^{[t/c_N]} - (I_d + c_N Q)^{[t/c_N]}|| = 0$$

En outre, on sait que $\lim_{N\to\infty}(I_d + \frac{1}{N}Q)^N = e^Q$. On peut en déduire que

$$\lim_{N\to\infty} P_N^{[t/c_N]} = \lim_{N\to\infty} (I_d + c_N Q)^{[t/c_N]} = \lim_{N\to\infty} ((I_d + [1/c_N]^{-1}Q)^{[1/c_N]})^t = e^{tQ} \quad \forall t \geq 0.$$

Ceci est équivalent à la convergence des marginales fini-dimensionnelles de $(R^N_{[t/c_N]})_{t\geq 0}$ vers celles de $(\mathcal{R}_t)_{t\geq 0}$: nous allons le démontrer dans ce qui suit, en montrant que $(e^{tQ})_{t\geq 0}$ est bien la famille de probabilités de transition qui caractérise le processus de sauts de matrice de sauts Q (dont on sait qu'il s'agit d'un Λ-coalescent, par définitionde ce dernier).
Commençons par un lemme général sur les processus de sauts, qui nous donne une définition alternative de la matrice de sauts :

Lemme 5. *Si $(X(t))_{t\geq 0}$ est un processus de sauts de matrice de sauts Q et d'ensemble d'états S fini, alors pour tous éléments $x \neq y$ de S on a*

$$\lim_{\substack{h\to 0\\ h>0}} \frac{1}{h}\mathbb{P}(X(h) = y | X(0) = x) = \lim_{\substack{h\to 0\\ h>0}} \frac{1}{h}\mathbb{P}_x(X(h) = y) = q(x, y).$$

Preuve : Si $h > 0$, en utilisant la propriété de Markov de $(X(t))_{t\in\mathbb{R}_+}$, on obtient l'égalité :

$$\begin{aligned}\frac{1}{h}\mathbb{P}(X(h) = y | X(0) = x) &= \frac{1}{h}\mathbb{P}_x(E_x^1 < h, X(E_x^1) = y, E_x^1 + E_y^1 > h)\\ &\quad + \frac{1}{h}\sum_{\substack{z\in S\\ z\neq y}} \mathbb{P}_x(X(E_x^1) = z, E_x^1 + E_z^1 \leq h, X(h) = y).\end{aligned}$$

Le premier terme du membre de droite correspond au cas où il n'y a qu'un saut entre 0 et h pour aller en y. Quant au second terme, il correspond aux cas où il y a au moins deux sauts avant d'arriver en y. Or, on a que :

$$\begin{aligned}\mathbb{P}(E_x^1 + E_z^1 \leq h) &= 1 - \int_{u+v\geq h, u\geq 0, v\geq 0} q_x e^{-q_x u} q_z e^{-q_z v} du dv\\ &= 1 - \frac{q_z e^{-q_x h} - q_x e^{-q_z h}}{q_z - q_x} \sim \frac{q_z q_x}{2} h^2 + o(h^2).\end{aligned}$$

On en déduit que le second terme du membre de droite de l'équation précédente est un $O(h^2)$ et que le premier terme est équivalent à

$$\frac{1}{h}\mathbb{P}_x(E_x^1 < h, X(E_x^1) = y) = \frac{1}{h}\int_0^h q_x e^{-q_x u} du \; p(x, y) \sim q_x p(x, y) = q(x, y).$$

Ceci termine la démonstration du lemme. □

On peut donc grâce à ce lemme faire l'approximation au voisinage de 0 pour $x \neq y \in S$

$$\mathbb{P}_x(X(h) = y) = q(x, y)h + o(h).$$

Ceci nous conduit au lemme suivant :

Lemme 6. (Equations de Chapman-Kolmogorov)
Si $(X(t))_{t\geq 0}$ est un processus de sauts de matrice de sauts Q et d'ensemble d'états S fini, alors en posant pour tous $x,\ y \in S$,

$$p_{x,y}(t) = \mathbb{P}_x(X(t) = y),$$

les fonctions ainsi définies vérifient les systèmes différentiels suivants :

$$\frac{d}{dt}p_{x,y}(t) = \sum_{z\in S} p_{x,z}(t)q(z,y) = \sum_{z\in S} q(x,z)p_{z,y}(t). \tag{1}$$

En notant $P(t) = (p_{x,y}(t);\ x,\ y \in S)$, la forme matricielle des équations précédentes est donnée par

$$\frac{d}{dt}P(t) = P(t)\cdot Q = Q\cdot P(t).$$

Preuve : Les deux systèmes d'équations différentielles s'obtiennent en utilisant la propriété de Markov, en décomposant l'intervalle de temps $[0, t+h]$ en $[0,t]\cup[t,t+h]$ pour les premières (équations dites «forward») ou en $[0,h]\cup[h,h+t]$ pour les secondes (équations dites «backward»).
Pour les équations «forward», on a pour $t \geq 0,\ x,\ y, \in S$ et $h > 0$,

$$\begin{aligned} p_{x,y}(t+h) &= \frac{\mathbb{P}(X(t+h) = y, X(0) = x)}{\mathbb{P}(X(0) = x)} \\ &= \sum_{z\in S} \mathbb{P}(X(t+h) = y, X(t) = z | X(0) = x). \end{aligned}$$

La propriété de Markov et celle d'homogénéité donnent successivement les identités

$$\begin{aligned} \mathbb{P}(X(t+h) = y, X(t) = z | X(0) = x) &= \mathbb{P}(X(t) = z | X(0) = x)\mathbb{P}(X(t+h) = y | X(t) = z, X(0) = x) \\ &= p_{x,z}(t)\mathbb{P}(X(t+h) = y | X(t) = z) \\ &= p_{x,z}(t)\mathbb{P}(X(h) = y | X(0) = z) \\ &= p_{x,z}(t)p_{z,y}(h). \end{aligned}$$

L'équation pour $p_{x,y}(t+h)$ devient

$$p_{x,y}(t+h) = \sum_{z\in S} p_{x,z}(t)p_{z,y}(h),$$

d'où

$$p_{x,y}(t+h) - p_{x,y}(t) = -p_{x,y}(t)(1 - p_{y,y}(h)) + \sum_{\substack{z\in S \\ z\neq y}} p_{x,z}(t)p_{z,y}(h).$$

Or d'après le lemme précédent, $p_{z,y}(h) = q(z,y)h + o(h)$ et comme l'ensemble S est fini,

$$1 - p_{y,y}(h) = \sum_{\substack{z\in S \\ z\neq y}} p_{z,y}(h) = \left(\sum_{\substack{z\in S \\ z\neq y}} q(z,y)\right)h + o(h) = -q(y,y)h + o(h).$$

En réinjectant ceci dans l'équation précédente, on obtient

$$\begin{aligned} p_{x,y}(t+h) - p_{x,y}(t) &= -p_{x,y}(t)\cdot(-q(y,y))h + \left(\sum_{\substack{z\in S \\ z\neq y}} p_{x,z}(t)q(z,y)\right)h + o(h) \\ &= \left(\sum_{z\in S} p_{x,z}(t)q(z,y)\right)h + o(h), \end{aligned}$$

ce qui permet de conclure.
La preuve se fait de manière similaire pour les équations «backward», l'équation de départ étant

$$\begin{aligned} p_{x,y}(t+h) &= \sum_{z\in S} \mathbb{P}(X(t+h)=y, X(h)=z|X(0)=x) \\ &= \sum_{z\in S} \mathbb{P}(X(h)=z|X(0)=x)\mathbb{P}(X(t+h)=y|X(h)=z) \\ &= \sum_{z\in S} p_{x,y}(h)p_{z,y}(t). \end{aligned}$$

□

Le lien avec le coalescent $(\mathcal{R}_t)_{t\geq 0}$ apparaît alors grâce au résultat suivant (on prendra X_t=$\mathcal{R}_t$) :

Lemme 7. (Représentation fonctionnelle de Chapman-Kolmogorov).
Si f est une fonction continue bornée sur S, alors

$$\frac{d}{dt}P(t,f) = P(t, Q\cdot f) = Q\cdot P(t,f),$$

où $(P(t,\cdot))$ est défini par

$$P(t,f)(x) = \mathbb{E}_x[f(X(t))]$$

pour tous $s\in S$ et f fonction continue bornée sur S. Sous leur forme fonctionnelle, ces équations différentielles s'expriment donc ainsi :

$$\frac{d}{dt}P(t) = Q\cdot P(t) = P(t)\circ Q. \tag{2}$$

Preuve : Si $x\in S$ et f est la fonction indicatrice de l'élément y, alors les équations (1) peuvent s'écrire sous la forme

$$\frac{d}{dt}\mathbb{E}_x[f(X(t))] = \mathbb{E}_x\left(\sum_{z\in S} q(X(t),z)f(z)\right) = \sum_{z\in S} q(x,z)\mathbb{E}_z[f(X(t))].$$

Puisque toute fonction f continue bornée sur S est la somme (finie si S est fini) d'une combinaison linéaire de telles fonctions, l'équation précédente s'étend donc à de telles fonctions. □

L'équation (2) est une équation différentielle de type exponentiel, cela suggère donc la solution formelle

$$P(t) = e^{tQ}.$$

On a donc bien, d'une part que la famille de matrices de transition du Λ-coalescent de matrice Q est $(e^{tQ})_{t\geq 0}$, et d'autre part que :

$$\lim_{N\to\infty} \mathbb{E}\Big[f(R^N_{[t/c_N]}|R^N_0=\xi)\Big] = \lim_{N\to\infty} P_N^{[t/c_N]}f(\xi) = e^{tQ}f(\xi) = P(t)f(\xi) = \mathbb{E}[f(\mathcal{R}_t)|\mathcal{R}_0=\xi]$$

pour toute fonction f continue bornée, ce qui conclut la preuve du théorème pour ce qui est des marginales uni-dimensionnelles. L'extension aux marginales de dimension finie s'obtient par récurrence en utilisant la propriété de Markov des processus.

4 Annexe

Démontrons la remarque (*) de la partie 1.2, que l'on rappelle :
(*) Pour le modèle de Wright-Fisher, on a

$$\begin{aligned} c_N &= \frac{1}{N}, \\ \mathbb{E}[(\nu_1)_k] &\to 1 \ quand \ N \to \infty, \\ \mathbb{E}[(\nu_1)_2(\nu_2)_2] &\to 1 \ quand \ N \to \infty. \end{aligned}$$

Preuve : Tout d'abord, comme $\nu_1 + \cdots + \nu_N = N$, on a $\mathbb{E}[\nu_1] = 1$.
Ensuite, $\mathbb{P}(\nu_1 = k) = \binom{N}{k}(\frac{1}{N})^k(\frac{N-1}{N})^{N-k}$.
En effet, la probabilité que $\nu_1 = k$ est égale à celle de choisir k individus parmi N, pour être les descendants du permier individu (chacun avec probabilité $\frac{1}{N}$), les $N-k$ restants étant des descendants de l'un des $N-1$ autres parents possibles (chacun avec probabilité $1-\frac{1}{N}$). On obtient donc :

$$\mathbb{P}(\nu_1 = k) = \binom{N}{k}\left(\frac{1}{N}\right)^k\left(\frac{N-1}{N}\right)^{N-k} = \frac{1}{k!}\frac{N(N-1)\dots(N-k+1)}{(N-1)^k}\left(1-\frac{1}{N}\right)^N,$$

d'où $\mathbb{P}(\nu_1 = k) \to \frac{e^{-1}}{k!}$ quand $N \to \infty$.
Ainsi, pour $\epsilon > 0$ et $n < N$, on a :

$$\begin{aligned} |\mathbb{E}[(\nu_1)_k] - 1| &= \Big|\sum_{i=k}^{N} i(i-1)\dots(i-k+1)\mathbb{P}(\nu_1 = i) - 1\Big| \\ &= \Big|\sum_{i=k}^{N} \frac{i(i-1)\dots(i-k+1)}{i!} i!\mathbb{P}(\nu_1 = i) - 1\Big| \\ &= \Big|\sum_{i=k}^{N} \frac{1}{(i-k)!} i!\mathbb{P}(\nu_1 = i) - e^{-1}\sum_{i=k}^{\infty}\frac{1}{(i-k)!}\Big| \\ &= \Big|\sum_{i=k}^{N} \frac{1}{(i-k)!}(i!\mathbb{P}(\nu_1 = i) - e^{-1})\Big| + \Big|e^{-1}\sum_{i=N+1}^{\infty}\frac{1}{(i-k)!}\Big| \\ &\le \Big|\sum_{i=k}^{n} \frac{1}{(i-k)!}(i!\mathbb{P}(\nu_1 = i) - e^{-1})\Big| + \Big|\sum_{i=n+1}^{N}\frac{2}{(i-k)!}\Big| + \Big|e^{-1}\sum_{i=N+1}^{\infty}\frac{1}{(i-k)!}\Big| \\ &\le 3\epsilon, \end{aligned}$$

si on prend n suffisament grand pour que $|\sum_{i=n+1}^{\infty}\frac{2}{(i-k)!}| < \epsilon$, puis N suffisament grand pour que $|\sum_{i=k}^{n}\frac{1}{(i-k)!}(i!\mathbb{P}(\nu_1 = i) - e^{-1})| < \epsilon$ et $|e^{-1}\sum_{i=N+1}^{\infty}\frac{1}{(i-k)!}| < \epsilon$.

Références

[Can74] C. Cannings. The latent roots of certain Markov chains arising in genetics : a new approach, I.Haploid models. *Adv. Appl. Probab.*, 6 :260–290, 1974.

[Kin82] J.F.C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13 :235–248, 1982.

[MS01] M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.*, 29 :1547–1562, 2001.

[Pit99] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27 :1870–1902, 1999.

[Sag03] S.Sagitov Convergence to the coalescent with simultaneous multiple mergers. J. Appli. Prob, 40 : 839-854, 2003.

[Sch00] J. Schweinsberg. Coalescents with simultaneous multiple collisions. Electron. J. Probab., 5 :1-50, 2000.

[Wak08] J. Wakeley. *Coalescent theory. An introduction.* Roberts & Company Publishers, 2008.