# Titre du rapport

Nom de l'auteur

31 octobre 2025

# 1 Introduction à la théorie de la coalescence.

## 1.1 Kingman coalescent est la limite de Wright-Fisher.

Nous disposons d'une population haploïde[1] de taille $N$. On commence à la génération $m = 0$ et nous souhaitons remonter le temps pour étudier la généalogie de $n < N$ individus, généralement on considère $n \ll N$. On appelle la lignée de $i \in [n] := [\![1, n]\!]$ la suite des ancêtres de $i$. Lorsque deux, ou plus, lignées choisissent le même parent à un instant donné, on dit qu'elles coalescent.

Formellement, on définit pour la génération $m \in \mathbb{N}$, la relation d'équivalence $\sim_m$ pour tout $i, j \in [n]$,

$$i \sim_m j \text{ ssi les lignées de } i \text{ et } j \text{ ont le même ancêtre à la génération m}$$

La classe de $i$ à la génération $m$, constituant un bloc, est,

$$C_m^{(N)}(i) := \{j \in [n], i \sim_m j\}$$

Puis, la famille des classes distinctes forme une partition de [n]

$$\Pi_m^{(N)} := \{C_m^{(N)}(i), i \in [n]\} \qquad K_m := |\Pi_m^{(N)}|$$

Dans le modèle historique de Wright-Fisher, à chaque génération, chaque individu choisit un parent au hasard dans la génération précédente, de manière indépendante et uniforme Fisher [1930], Wright [1931]. On a ainsi que $(\Pi_m^{(N)})_{m\geq 0}$ est une chaîne de Markov homogène à temps discret sur les partitions de $[n]$.

Notons $(\cdot)_k = \prod_{i=\cdot-k+1}^{\cdot} i$. On a que la probabilité que $k \geq 2$ lignées de $b$ blocs coalescent en une génération est donnée par,

$$\binom{b}{k}\frac{(N)_{b-k+1}}{N^b} = \begin{cases} \binom{b}{2}\frac{1}{N} + \mathcal{O}(\frac{1}{N^2}) & \text{si } k = 2 \\ \mathcal{O}(\frac{1}{N^2}) & \text{si } b \geq k \geq 3 \\ 0 & \text{sinon} \end{cases}$$

On remarque donc que l'évènement d'une coalescence est rare lorsque $N$ est grand. On considère ainsi le processus de Markov $(\Pi^{(N)}(t))_{t\geq 0}$ défini par,

$$\Pi^{(N)} : t \in \mathbb{R}^+ \longmapsto \Pi_{\lfloor tN \rfloor}^{(N)}$$

Notons le semi-groupe associé

$$P_t^N : (x, y) \mapsto \mathbb{P}(\Pi^{(N)}(t) = y | \Pi^{(N)}(0) = x)$$

Considérons $x$ une partition de $[n]$ avec $k$ blocs, $y$ une partition de $[n]$, avec $k - 1$ blocs, obtenue en fusionnant deux blocs de $x$, et $y^+$ une partition de $[n]$ en fusionnant plus de deux blocs de $x$. Puisque $t \mapsto P_t^N$ est en escalier, le générateur infinitésimal est donné par,

$$q^{(N)} : (x, z) \longmapsto N(P_{1/N}^N(x, z) - \delta_{x,z}) = \begin{cases} N\left(\frac{1}{N} + \mathcal{O}(\frac{1}{N^2}) - 0\right) = 1 + \mathcal{O}(\frac{1}{N}) & \text{si } z = y \\ N\left(\mathcal{O}(\frac{1}{N^2}) - 0\right) = \mathcal{O}(\frac{1}{N}) & \text{si } z = y^+ \\ N\left(1 - \binom{k}{2}\frac{1}{N} + \mathcal{O}(\frac{1}{N^2}) - 1\right) = -\binom{k}{2} + \mathcal{O}(\frac{1}{N}) & \text{si } z = x \end{cases}$$

1. Chaque individu porte une seule copie du génome, modélisée par un unique parent.

En passant à la limite sur $N$ vers l'infini, nous obtenons le coalescent de Kingman Kingman [1982]. Son générateur est donné par,

$$q : (x, z) \longmapsto \lim_{N\to\infty} q^{(N)}(x, z) = \begin{cases} 1 & \text{si } z = y \\ -\binom{k}{2} & \text{si } z = x \\ 0 & \text{sinon} \end{cases}$$

Ce modèle considère uniquement le fait que deux lignées peuvent coalescer à la fois.

### 1.2 Λ-Coalescent.

Intéressons nous à un cas plus général où un certain nombre de lignées peuvent coalescer simultanément. Nous restons dans l'esprit d'un modèle limite où $N$ tend vers l'infini afin de considérer une proportion $x \in [0, 1]$ de la population descendant d'un même individu. La description du générateur infinitésimal de ce processus de Markov repose sur le théorème suivant,

**Théorème 1** (Pitman-Sagitov Pitman [1999], Sagitov [1999])**.** *Il existe un processus de Markov, appelé $\Lambda$-coalescent, échangeable à collisions multiples simples si et seulement s'il existe une mesure finie $\Lambda$ sur $[0, 1]$ telle que, lorsqu'on a $b$ blocs, pour tout $2 \leq k \leq b$ le taux auquel chaque $k$-uplet fixé de blocs fusionne vaut,*

$$\lambda_{b,k} = \int_0^1 x^{k-2}(1-x)^{b-k}\, \Lambda(dx)$$

Sous des hypothèses raisonnables telles que l'absence de mémoire (processus de Markov), la permutation invariante des blocs (échangeabilité) et qu'à chaque instant il ne peut pas y avoir plusieurs coalescences n'ayant pas le même ancêtre (collisions multiples simples), ce résultat montre que la dynamique d'un processus de coalescence multiple est entièrement caractérisée par une mesure finie $\Lambda$ sur $[0, 1]$.

Dans ce rapport nous simulons différents $\Lambda$-coalescents pour différentes mesures $\Lambda$ et TODO : dire ce qu'on veut faire réellement que juste faire des plots.

## 2 ? ?

Je sais pas si on fait une "liste" de mesure qu'on regarde, je pensais plutôt renrer dans le vif du sujet avec la problémtique définie et au fur et à mesure on montre les différents plot avec les différentes mesures testées.

Parler de Lambda $= \delta_0$ : Kingman (dire que la theorie "généralise") + Parler de beta ?

TODO :

- Image vulgarisatrice de ce qu'on fait (j'ai une idée donc à voir si c'est utile)
- Détails des calculs d'Axcel pour différentes mesure : nouvelle section + subsections
- Enchainer avec une succession de beaux plots
- Lancer une problématique (à en discuter voir où on veut aller : trouver un paramètre par simulation (temps d'un unique ancêtre, pb de coalescence multiple, . . .), estimer la distance entre deux modèles (KS, Wasserstein, . . .), usage détourné de la coalescence pour faire quelque chose d'atypique)

## Références


Ronald A. Fisher. *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford, 1930.

Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2) :97–159, 1931.

J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3) :235–248, 1982. doi : 10.1016/0304-4149(82)90011-4.

Jim Pitman. Coalescents with multiple collisions. *The Annals of Probability*, 27(4) :1870–1902, 1999.

Serik Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *Journal of Applied Probability*, 36(4) :1116–1125, 1999.