Modélisation mathématique du plus récent ancêtre commun d'une population

Rémy Mahfouf
Sous la direction d'Amandine Veber

1 Introduction

Les biologistes essayent de trouver une filiation commune à l'humanité actuelle depuis de nombreuses années. On peut se demander si tous les êtres humains vivant actuellement sur terre ont un ancêtre commun, et si oui combien de générations il est nécessaire de remonter pour trouver cet ancêtre. Ce mémoire répond à une question analogue en présentant différents modèles de reproduction. Cependant on peut d'ores et déjà dire qu'il faut rester réservé sur la cohérence avec les résultats empiriques que les biologistes constatent : en effet on va proposer dans la suite de ce papier des modèles de reproduction qui mettent en jeu un aléa de reproduction avec un choix uniforme de deux parents dans la génération précédente. Ceci occulte complètement l'influence de la géographie, du temps de survie, des facteurs socio-économiques et culturels qui impactent énormément sur l'appareillement des parents (si on regarde par exemple des être humains). La taille constante de la population est elle aussi assez éloignée de la réalité. Ce mémoire n'a donc pas pour but d'expliquer l'arbre généalogique de l'humanité actuelle mais plutôt de proposer des modèles mathématiques ainsi que des simulations numériques qui peuvent servir de base de travail à certaines modélisations futures des biologistes pour des espèces précises. La littérature déjà disponible sur ce sujet repose essentiellement sur le modèle de Wright-Fisher qui suppose le choix pour chaque individu d'un unique parent dans la génération précédente. Les principaux résultats sur le sujets sont détaillés dans [3]. Ce choix est fait de manière aléatoire, uniforme et indépendant de celui des autres individus de la génération. Dans le modèle de Wright-Fisher, on se donne un entier naturel n qui représente la taille de la population et qui sera constant au cours du processus. Dans le cadre du modèle de Wright-Fisher, on peut répondre par l'affirmative à la question initiale : avec une probabilité qui tend vers 1 lorsque $n \to \infty$, il existe un ancêtre commun à tous les individus. De plus on sait montrer qu'il faut remonter environ $2n$ générations par rapport à la génération actuelle pour trouver le plus récent ancêtre commun.
On s'intéresse dans ce papier au modèle de reproduction suivant:

- On se donne une population initiale de taille n fixée. Cette taille restera fixée au cours du temps.
- A la génération $t+1$, chaque individu choisit de manière indépendante et uniforme deux parents de la génération t.

- Il est possible qu'un individu ait les deux mêmes parents si le choix aléatoire de ses deux parents est le même.

- Tous les individus de la génération t meurent au début de la génération $t+1$, juste après le choix de la filiation. En particulier ils ne sont pas considérés comme des individus de la génération $t+1$.

On note $\mathcal{T}_n$ le nombre de générations qu'il faut remonter au minimum dans la généalogie pour trouver un ancêtre commun à tous les individus de la population actuelle. Les deux résultats principaux de la première partie de ce mémoire sont résumés comme suit:
Premièrement $\lim\limits_{n\to\infty} \frac{\mathcal{T}_n}{\log(n)} = 1$ en probabilités, où log désigne le logarithme en base 2. De plus avec une probabilité qui tend vers 1 lorsque $n \to \infty$, si on remonte environ $1,77\log(n)$ générations alors chaque individu est soit un ancêtre de toute la population actuelle, soit n'est ancêtre de personne. On notera dans la suite l'abréviation AC pour un ancêtre commun de la génération actuelle (et donc aux suivantes) et $PRAC$ pour un des plus récent ancêtre commun (il n'est pas à priori unique) de la génération actuelle.
Ces résultats sont cohérents avec les conclusions du modèle Wright-Fisher que l'on vient d'évoquer. En effet on utilise dans le second modèle une généalogie double, c'est à dire avec deux parents potentiellement différents. Il est donc naturel de penser qu'avec un doublement du nombre de parents, on pourra à nouveau trouver un ancêtre commun et celui-ci apparaîtra plus rapidement dans la généalogie.
On énonce à présent de manière plus formelle les deux résultats que l'on veut démontrer dans la première partie de ce mémoire.

Théorème 1 *Soit $\mathcal{T}_n$ le nombre de générations qu'il faut remonter dans la généalogie pour trouver un $PRAC$. Alors $\lim\limits_{n\to\infty} \frac{\mathcal{T}_n}{\log(n)} = 1$ en probabilités.*

Théorème 2 *Soit $\mathcal{U}_n$ le nombre de générations qu'il faut remonter dans la généalogie pour que chaque individu vérifie une des assertions suivantes :*

- *Il est ancêtre commun de toute la population actuelle.*

- *Il n'est ancêtre d'aucun individu de la population actuelle.*

Soit γ la plus petite des solutions $\gamma e^{-\gamma} = 2e^{-2}$ et $\xi = \frac{-1}{\log(\gamma)} \approx 0,77$.
Alors $\lim\limits_{n\to\infty} \frac{\mathcal{U}_n}{(1+\xi)\log(n)} = 1$ en probabilités.

Je tiens à remercier chaleureusement Amandine Veber, mon encadrante pour ce travail, de m'avoir proposé ce sujet intéressant. Cette dernière a été d'une grande aide par ses explications, pistes de recherche et multiples relectures. La réalisation de ce mémoire lui doit beaucoup.

2 Résultats préparatoires

Dans le but de prouver les deux théorèmes énoncés dans l'introduction, nous allons à présent démontrer plusieurs lemmes et résultats généraux, pour ensuite mettre à profit les particularités du modèle. Pour des raisons de lisibilité et de simplicité des notations, nous prenons la convention de décrire les générations successives depuis la génération zéro.

A la génération t, la population est constituée de n individus (on rappelle que la taille de la population ne varie pas au cours du temps). A la génération t, les individus sont notés $I_{t,1}, I_{t,2}, ...I_{t,n}$. La numérotation des individus de la génération t est ici totalement arbitraire et n'aura pas d'incidence sur les résultats. On choisit aussi $\mu_{t,1}, \nu_{t,1}, \mu_{t,2}, \nu_{t,2}...\mu_{t,n}, \nu_{t,n}$ des variables aléatoires indépendantes identiquement distribuées uniformément sur $[\![1, n]\!]$. On interprète $\mu_{t,j}$ et $\nu_{t,j}$ comme les parents de l'individu $I_{t,j}$ c'est à dire que $I_{t,j}$ est descendant direct de $I_{t-1,\mu_{t,j}}$ et de $I_{t-1,\nu_{t,j}}$. On rappelle enfin que les deux parents peuvent être les mêmes. On peut à présent définir par récurrence sur l'entier t (qui correspond à la génération) la suite d'ensembles aléatoires $\mathcal{G}_0^i, \mathcal{G}_1^i \subseteq [\![1, n]\!]$ comme suit. On pose $\mathcal{G}_0^i = \{i\}$ et pour $t \geq 1$ entier on définit l'ensemble $\mathcal{G}_t^i = \left\{j \leq n : \mu_{t,j} \in \mathcal{G}_{t-1}^i \text{ ou } \nu_{t,j} \in \mathcal{G}_{t-1}^i\right\}$. L'ensemble $\mathcal{G}_t^i$ représente les descendants à la génération t de l'individu $I_{i,0}$. On note aussi G_t^i le cardinal de $\mathcal{G}_t^i$. Il est clair que les ensembles $\mathcal{G}_t^i$ et les entiers G_t^i sont mesurables. La numérotation des individus à la génération t étant arbitraire, il est naturel de faire porter l'étude sur la suite des cardinaux G_t^i. On obtient une première proposition sur la distribution de la variable aléatoire G_t^i.

Proposition 1 *Le processus aléatoire $(G_t)_{t\geq 0}$ est une chaîne de Markov de noyau de transition Q définit par*

$$Q(r, s) = \binom{n}{s} p(r)^s (1 - p(r))^{n-s} \qquad p(r) = \tfrac{2r}{n} - \tfrac{r^2}{n^2}$$

Ce que l'on peut résumer par : $(G_{t+1}^i | G_t^i) \sim \mathrm{Bin}(n, \frac{2G_t^i}{n} - \frac{(G_t^i)^2}{n^2})$

<u>**Preuve**</u>

- Les variables aléatoires $\mu_{t+1,j}$ et $\nu_{t+1,j}$ sont indépendantes. Les parents d'un individu sont choisis de manière indépendante et uniforme dans la génération précédente.

- La probabilité conditionnelle que l'individu $I_{t+1,j}$ ait au moins un parent parmi les G_t^i individus de $\mathcal{G}_t^i$ connaissant G_t^i est :
$\mathbb{P}(\left\{\mu_{t+1,j} \in \mathcal{G}_{t-1}^i\right\} \cup \left\{\mu_{t+1,j} \in \mathcal{G}_{t-1}^i\right\} | G_t^i) = \frac{G_t^i}{n} + \frac{G_t^i}{n} - (\frac{G_t^i}{n})^2$ par indépendance des événements $\left\{\mu_{t+1,j} \in \mathcal{G}_{t-1}^i\right\}$ et $\left\{\mu_{t+1,j} \in \mathcal{G}_{t-1}^i\right\}$.

- Le noyau de transition de G_t^i à G_{t+1}^i est indépendant de la génération t donc on peut voir le processus $(G_t^i)_{t\geq 0}$ comme une chaîne de Markov homogène en temps, de matrice de transition Q annoncée. $\diamond$

Dans la suite de notre démonstration, on réalisera une approximation du processus $(G_t)_{t\geq 0}$ par un processus de Galton-Watson avec une loi de reproduction de Poisson de paramètre 2. On rappelle ici la définition d'un processus de Galton-Watson.

Définition 1 *Soit μ une mesure de probabilité sur $\mathbb{N}$ et $l \in \mathbb{N}$. On commence par se donner des variables aléatoire $(\xi_{t,j})_{(t,j)\in\mathbb{N}\times\mathbb{N}^*}$, mutuellement indépendantes, identiquement distribuées de loi μ. On définit par récurrence la suite de variables aléatoires $(X_n)_{n\geq 0}$ par :*

$$X_0 = l; \quad X_{n+1} = \sum_{i=1}^{X_n} \xi_{n,j}$$

Le processus $(X_n)_{n\geq 0}$ est un processus de Galton-Watson de valeur initiale l et de loi de reproduction μ.

Les processus de Galton-Watson sont tout à fait adaptés à l'étude de l'évolution temporelle de la taille d'une population en supposant des reproductions indépendantes suivant la loi μ. Bien qu'il s'agisse ici d'une étude a taille de population constante, on peut tout de même exploiter (après approximation bien choisie) dans notre étude des résultats classiques sur les processus de Galton-Watson.

Lemme 1 *Soit $(Y_t)_{t\geq 0}$ un processus de Galton-Watson avec une loi de reproduction de Poisson de paramètre 2. On suppose que $Y_0 = G_0 = 1$.*
On définit les variables $\tau_b^Y = \inf\{t : Y_t \geq b\}$, $\tau_{0,b}^Y = \inf\{t : Y_t \geq b \text{ ou } Y_t = 0\}$, $\tau_b^G = \inf\{t : G_t \geq b\}$ et $\tau_{0,b}^G = \inf\{t : G_t \geq b \text{ ou } G_t = 0\}$. Soient m et b entiers vérifiant $mb^2 = o(n)$ lorsque $n \to \infty$. Asymptotiquement on a:
$\mathbb{P}(\tau_b^G > m) = \mathbb{P}(\tau_b^Y > m)(1 + o(1))$ et $\mathbb{P}(\tau_{0,b}^G > m) = \mathbb{P}(\tau_{0,b}^Y > m)(1 + o(1))$

Preuve

- On commence la preuve par un calcul de rapport de vraisemblance sur les lois de transitions. On a pour $x, y \in \mathbb{N}$:
  $L(y|x) = \frac{\mathbb{P}(G_{t+1}=y|G_t=x)}{\mathbb{P}(Y_{t+1}=y|Y_t=x)} = \frac{\mathbb{P}(\mathrm{Bin}(n,\frac{2x}{n}-\frac{x^2}{n^2})=y)}{\mathbb{P}(Poisson(2x)=y)} = \frac{\binom{n}{y}(\frac{2x}{n}-\frac{x^2}{n^2})^y(1-\frac{2x}{n}+\frac{x^2}{n^2})^{n-y}}{e^{-2x}\frac{(2x)^y}{y!}}$.

- En simplifiant $L(y|x) \leq e^{2x}(1 - \frac{2x}{n} + \frac{x^2}{n^2})^{n-y}$. On passe alors au log et $\log(L(y|x)) \leq 2x + (n-y)(\frac{2x}{n} - \frac{x^2}{n^2}) \leq \frac{x^2+2xy}{n}$.

- Si $x < b$ et $y < b$ on a alors $\log(L(y|x)) \leq \frac{3b^2}{n}$.

- On obtient de même une borne inférieure $\log(L(y|x)) \geq \frac{-5b^2}{2n}(1 + \mathcal{O}(\frac{b}{n}))$ et donc $\log(L(y|x)) \geq \frac{-3b^2}{n}$ pour n assez grand.

- Si on fixe $x_1, x_2 \ldots , x_m \leq b$, la propriété de Markov de la chaîne donne :
  $\mathbb{P}(G_1 = x_1, G_2 = x_2 \ldots, G_m = x_m) = \mathbb{P}(G_0 = 1)Q(1, x_1)Q(x_2)Q(x_{m-1}, x_m)$
  puis $\mathbb{P}(G_1 = x_1, ..., G_m = x_m) = \mathbb{P}(Y_1 = x_1, .., Y_m = x_m)L(x_1|1)...L(x_m|x_{m-1})$.

- On utilise la majoration sur les vraisemblances et on obtient :
  $\mathbb{P}(G_1 = x_1, ..., G_m = x_m) \leq \mathbb{P}(Y_1 = x_1, ..., Y_m = x_m)e^{\frac{3mb^2}{n}}$.

- Le même raisonnement utilisant la minoration montre
$\mathbb{P}(G_1 = x_1, ..., G_m = x_m) \geq \mathbb{P}(Y_1 = x_1, ..., Y_m = x_m)e^{\frac{-3mb^2}{n}}$.

- Par définition : $\mathbb{P}(\tau_b^G > m) = \sum_{0 \leq x_1, x_2, ..., x_m < b} \mathbb{P}(G_1 = x_1, ..., G_m = x_m)$.

- On obtient en appliquant le résultat précédent à chaque $m - uplet$:
$$\begin{aligned}\mathbb{P}(\tau_b^G > m) &= \sum_{0 \leq x_1, x_2, ..., x_m < b} \mathbb{P}(G_1 = x_1, G_2 = x_2, ..., G_m = x_m) \\ &\leq \sum_{0 \leq x_1, x_2, ..., x_m < b} \mathbb{P}(Y_1 = x_1, Y_2 = x_2, ..., Y_m = x_m)e^{\frac{3mb^2}{n}} \\ &= \mathbb{P}(\tau_b^Y > m)e^{\frac{3mb^2}{n}}\end{aligned}$$

- De même pour la minoration $\mathbb{P}(\tau_b^G > m) \geq \mathbb{P}(\tau_b^Y > m)e^{\frac{-3mb^2}{n}}$. Comme $mb^2 = o(n)$ on obtient le premier résultat annoncé.

- Pour le second résultat, la preuve est exactement la même jusqu'à la sommation. La différence provient du fait qu'on doit interdire aux x_i de prendre la valeur 0. La nouvelle somme est :
$$\begin{aligned}\mathbb{P}(\tau_{0,b}^G > m) &= \sum_{0 < x_1, x_2, ..., x_m < b} \mathbb{P}(G_1 = x_1, G_2 = x_2, ..., G_m = x_m) \\ &\leq \sum_{0 < x_1, x_2, ..., x_m < b} \mathbb{P}(Y_1 = x_1, Y_2 = x_2, ..., Y_m = x_m)e^{\frac{3mb^2}{n}} \\ &= \mathbb{P}(\tau_{0,b}^Y > m)e^{\frac{3mb^2}{n}}\end{aligned}$$

- On conclut que $\mathbb{P}(\tau_{0,b}^G > m) \geq \mathbb{P}(\tau_{0,b}^Y > m)e^{\frac{-3mb^2}{n}}$. $\diamond$

Avec cette approximation, on peut maintenant espérer utiliser des résultats classiques sur sur les processus de Galton-Watson. C'est ce que l'on réalise maintenant.

Lemme 2 *Soit $(Y_t)_{t\geq 0}$ un processus de Galton-Watson avec $Y_0 = 1$ et une loi de reproduction de Poisson de paramètre 2. On définit la fonction génératrice $\psi(z) = \mathbb{E}(z^{Y_1}) = e^{-2+2z}$ et on pose $\rho = \mathbb{P}(\exists t, Y_t = 0)$ la probabilité d'extinction du processus. Alors ρ est la plus petite des deux solutions de l'équation $\psi(\rho) = \rho$ et vaut approximativement $\rho \approx 0.20319$. De plus $\rho = \frac{\gamma}{2}$ où γ est définie dans 2 . Enfin si on note $\psi_t = \psi \circ \psi \circ ...$ la composée t-ème de ψ, alors pour $0 \leq z \leq \rho$ $\psi_t(z) \to \rho$ lorsque $t \to \infty$.*

Pour la suite de notre introduction préparatoire on pose $g_t = \frac{G_t}{n}$ la proportion de descendants de $I_{0,1}$ à la génération t. Le calcul d'espérance pour une loi Binomiale donne $\mathbb{E}(g_{t+1}|g_t) = g_t(2 - g_t)$.
Pour les premières générations (ce qui correspond à un g_t petit), la viabilité de la descendance est loin d'être assurée. La probabilité d'extinction est raisonnable sans être très forte. C'est dans cette partie de la généalogie qu'il est intéressant d'approcher notre processus $(G_t)_{t\geq 0}$ par un processus Galton-Watson avec loi de reproduction de Poisson de paramètre 2.

On termine la présentation des résultats introductifs par l'inégalité de Bernstein qui permet d'évaluer la déviation d'une loi Binomiale par rapport à sa moyenne et montre que si G_t est suffisamment grand, alors $g_{t+1} \approx 2g_t - g_t^2$.

Proposition 2 *Soit* $X \sim \mathrm{Bin}(n,p)$ *une loi Binomiale et* $r > 0$ *fixé. Alors :*
$\mathbb{P}(X \geq np + r) \leq \exp(\frac{-r^2}{2np(1-p)+\frac{2r}{3}})$ *et* $\mathbb{P}(X \leq np + r) \leq \exp(\frac{-r^2}{2np(1-p)+\frac{2r}{3}})$

Ces inégalités proviennent de l'inégalité de Hoeffding, appliquée judicieusement.

3 Preuve du Théorème 1

La preuve du premier théorème est relativement longue. Pour bien comprendre le déroulement général du raisonnement, on commence par en donner les grandes étapes. On trouve en premier lieu un potentiel ancêtre commun, puis on évalue le temps nécessaire pour que sa descendance franchisse différents paliers(que l'on détaillera par la suite) jusqu'à devenir ancêtre commun de tous les individus de la génération considérée.

1. On commence par la génération 0 puis on avance dans le temps comme convenu dans l'introduction. A la fin de cette première étape on identifie un individu I de la génération 0 dont le nombre de descendants est petit comparé à n mais assez grand pour qu'il y ait une faible probabilité que sa descendance s'éteigne. On va chercher une génération t où l'individu I a au moins $\log(n)^2$ descendants. Avec une probabilité qui tend vers 1 lorsque n tend vers l'infini, cela arrive au bout d'un temps $t = o(\log(n))$, négligeable devant $\log(n)$. Ce résultat découle de l'approximation par le processus de Galton-Watson.
   Le reste de la preuve consiste à démontrer que l'individu I que l'on vient de choisir devient un ancêtre commun à tous les individus après au plus $(1+\varepsilon)\log(n)$ générations, et ceci pour tout $\varepsilon > 0$.

2. Soit $\beta \in]0;1[$ fixé. On va chercher dans cette partie une génération t ou l'individu I possède au moins n^β descendants. Comme n^β est négligeable devant n, la relation $\mathbb{E}(g_{t+1}|g_t) = g_t(2 - g_t)$ montre que le nombre de descendants de I double environ à génération. Cette phase ne durera au plus que $\log(n^\beta) = \beta \log(n)$ générations.

3. Dans cette étape on montre que le nombre de descendants de l'individu I passe ensuite de n^β à $\frac{n}{2}$ avec une probabilité qui tend vers 1 lorsque $n \to \infty$. Le facteur multiplicatif entre les générations successives est d'au moins $2 - \frac{1}{2} = \frac{3}{2}$. Il sera donc nécessaire d'attendre un temps de l'ordre $\log_{3/2}(\frac{n/2}{n^\beta})$ générations. Si on choisit β suffisamment proche de 1, on peut faire cette étape en un temps arbitrairement petit devant $\log(n)$

4. On va maintenant regarder l'ensemble des individus qui ne sont pas descendants de I. On s'attend à ce que la proportion $B_t = 1 - g_t$ soit élevée au carré à chaque nouvelle génération, ce qui impose une décroissance rapide vers 0 du nombre de

non descendants de I.
Si on fixe $\frac{1}{2} < \alpha < \frac{2}{3}$, le temps nécessaire pour faire passer B_t de $\frac{1}{2}$ à $\frac{1}{n^\alpha}$ est de l'ordre de $\log(\log(n))$ générations.

5. La rechercher de l'AC se termine lorsque B_t vaut exactement 0 et alors I est devient un ancêtre commun. Ceci arrivera en une seule génération avec une probabilité qui tend vers 1 lorsque n tend vers l'infini.

6. Pour la borne supérieure on recombine les résultats précédents pour montrer que pour ε fixé on a $\mathbb{P}(\tau_n \leq (1+\varepsilon)\log(n))$ tend vers 1 pour n qui tend vers l'infini.

7. Pour la borne inférieure, on utilisera l'inégalité de Bernstein.

3.1 Étape 1

On montre dans ce paragraphe qu'avec une probabilité qui tend vers 1 lorsque $n \to \infty$, on peut trouver une génération $t = o(\log(n))$, tel qu'au moins $\log^2(n)$ individus de la génération t ont un ancêtre en commun.

Lemme 3 *Soit* $\tau_b = \inf\{t : Y_t \geq b\}$. *On suppose que* $Y_0 = G_0 = 1$. *Alors* $\limsup_{n\to\infty} \mathbb{P}\big(\tau_{\log(n)^2} \leq 3\log(\log(n))\big) > 0$.

Preuve

- On note $b = \log^2(n)$ et $m = [3\log(\log(n))]+1$. Soit $(Y_t)_{t\geq 0}$ un processus de Galton-Watson avec une loi de reproduction de Poisson de paramètre 2. On pose $M_t = \frac{Y_t}{2^t}$. Alors $(M_t)_{t\geq 0}$ est une martingale pour la filtration canonique.

- En effet $\mathbb{E}(M_{t+1}|M_t) = \frac{1}{2^{t+1}}\mathbb{E}(\sum_{i=1}^{Y_t} \xi_{t,j}|M_t)$. Par indépendance de M_t vis à vis des $(\xi_{k,j})_{j\geq 1}$, $\mathbb{E}(M_{t+1}|M_t) = \frac{1}{2^{t+1}} \sum_{i=1}^{Y_t} \mathbb{E}(\xi_{t,j}) = \frac{2Y_t}{2^{t+1}} = M_t$.
De plus M_t est bien intégrable, donc $(M_t)_{t\geq 0}$ est bien une martingale positive bornée dans $\mathcal{L}^1$.

- La suite $(M_t)_{t\geq 0}$ converge donc presque sûrement vers une limite M_∞.
De plus $\lim_{t\to\infty} \mathbb{P}(Y_t = 0) = \rho$.

- On remarque que $\mathbb{P}(\tau_b^Y > m) \leq \mathbb{P}(Y_m < b) = \mathbb{P}(M_m < 2^{-m}b)$.
On applique le lemme de Fatou, comme $2^{-m}b \to 0$ lorsque $n \to \infty$
$\limsup_{n\to\infty} \mathbb{P}\big(\tau_b^Y > m\big) \leq \limsup_{n\to\infty} \mathbb{P}\big(M_m < b2^{-m}\big) \leq \mathbb{P}(\limsup_{n\to\infty} \{M_m < b2^{-m}\})$.
On en déduit :

- $$\begin{aligned} \mathbb{P}(\limsup_{n\to\infty} \{M_m < b2^{-m}\}) &= \mathbb{P}(M_m < b2^{-m} \text{ une infinité de fois }) \\ &= \mathbb{P}(M_\infty = 0) < \quad 1 \end{aligned}$$

- Par le lemme 1, $\mathbb{P}(\tau_b > m) = \mathbb{P}(\tau_b^Y)(1 + o(1))$ et on peut conclure $\limsup_{n\to\infty} \mathbb{P}\big(\tau_b > m\big) \leq \rho < 1$

- Finalement $\liminf_{n\to\infty} \mathbb{P}\big(\tau_b \leq m\big) \geq 1 - \rho > 0$ $\diamond$

Proposition 3 *On définit G_t^i le nombre de descendants à l'instant t de l'individu $I_{0,i}$. On définit $G_t^* = \max_{i\leq n}\left\{G_t^i\right\}$ et $\tau_b^{G^*} = \inf\left\{t : G_t^* \geq b\right\}$. Alors $\lim_{n\to\infty} \mathbb{P}(\left\{\tau_{\log^2(n)}^{G^*} = o(\log(n))\right\}) =$ 1.*

<u>**Preuve**</u>

- On pose $m_n = [3\log(\log(n))]$ et on prend $(k_n)_{n\geq 0}$ tel que $k_n \to \infty$ et $k_n m_n = o(\log(n))$. On procède par essais successifs.

- On commence par prendre $I_{0,1}$ et on regarde sa descendance après m_n générations. On dit que l'essai est un succès si à la génération m_n l'individu $I_{0,1}$ a une descendance d'au moins $\log^2(n)$ individus.

- Le lemme (3) nous assure que la probabilité d'un succès est strictement positive, supérieure à $c > 0$.

- Si l'essai est un échec, on fait un nouvel essai avec l'individu $I_{m_n,1}$ et on regarde si sa descendance compte au moins $\log^2(n)$ individus après m_n nouvelles générations.

- On recommence l'expérience jusqu'au premier succès.

- La probabilité que l'on ait constamment échoué jusqu'à la génération $k_n m_n$ vaut au plus $(1 - c)^{k_n} \to 0$ lorsque $n \to \infty$.

- On en déduit qu'avec une probabilité qui tend vers 1, on peut trouver un entier $\kappa \leq k_n - 1$ tel que l'individu $I_{\kappa m_n,1}$ possède au moins $\log^2(n)$ descendants à la génération $(1 + \kappa)m_n$.

- On note I un ancêtre de la génération 0 de $I_{\kappa m_n,1}$. $\diamond$

On va montrer maintenant que pour $\varepsilon > 0$, l'individu I que l'on vient d'identifier est un ancêtre commun de toute la population actuelle après $(1 + \varepsilon)\log(n)$ générations avec une probabilité qui tend vers 1 lorsque $n \to \infty$.

3.2 Étape 2

Une application de l'inégalité de Bernstein nous donne le lemme suivant.

Lemme 4 *On fixe $\delta \leq \frac{3}{4}$ et on se donne $t \in \mathbb{N}$ tel que $G_t \leq \frac{\delta n}{20}$.*
Alors $\mathbb{P}(G_{t+1} \leq (2 - \delta)G_t | G_t) \leq \exp(-\frac{\delta^2 G_t}{5})$.

Nous allons à présent montrer que la probabilité que la seconde étape décrite dans le plan de la preuve dure plus de $\log(n)$ générations tend vers 0 lorsque $n \to \infty$.

Proposition 4 *On suppose que* $G_0 \geq \log^2(n)$ *et on fixe* $0 < \beta < 1$. *Soit* $T_2 = \inf\{t : G_t \geq n^\beta\}$. *Alors* $\mathbb{P}(T_2 > \log(n)) = o(\frac{1}{n})$ *lorsque* $n \to \infty$.

Preuve

- On fixe $0 < \delta < \frac{3}{4}$ vérifiant $\log(2-\delta) > \beta$ puis on définit l'entier $b(n) = [\log_{2-\delta}(\frac{n^\beta}{\log^2(n)})]$. Pour $n \geq 3$, $b(n) \leq \frac{\beta \log(n)}{\log(2-\delta)} \leq \log(n)$.
- Par croissance des événements $\mathbb{P}(T_2 > \log(n)) \leq \mathbb{P}(T_2 > b(n))$.
  Il suffit donc de montrer que $\mathbb{P}(T_2 > b(n)) = o(\frac{1}{n})$ pour obtenir le résultat désiré.
- L'inégalité $\{T_2 > b(n)\}$ impose qu'il existe au moins un entier $t \leq b(n)$ qui vérifie $G_{t+1} < (2-\delta)G_t$. Le plus petit des entiers t vérifiant cette condition vérifie aussi que $G_t \geq \log^2(n)$.
- La croissance puis la sous-additivité donnent :
$$\begin{aligned}\mathbb{P}(T_2 > b(n)) &\leq \mathbb{P}\Big(\bigcup_{t \leq b(n)-1} \big\{G_{t+1} < (2-\delta)G_t, G_t \geq \log^2(n), T_2 > b(n)\big\}\Big)\\ &\leq \sum_{t=0}^{b(n)-1} \mathbb{P}(\big\{G_{t+1} < (2-\delta)G_t, \log^2(n) \leq G_t \leq n^\beta\big\})\end{aligned}$$
- Pour n suffisamment grand, $n^\beta \leq \frac{\delta n}{20}$ et on peut appliquer le lemme (4), qui assure $\mathbb{P}(G_{t+1} \leq (2-\delta)G_t | G_t) \leq \exp(-\frac{\delta^2 G_t}{5})$.
- En combinant ces résultats $\mathbb{P}(\{T_2 > b(n)\}) \leq b(n) n^{-\frac{\delta^2 log(n)}{5}} = o(\frac{1}{n})$. $\diamond$

3.3 Etape 3

Cette étape commence à une génération où l'individu I possède au moins n^β descendants et se termine quand le nombre de descendants dépasse $\frac{n}{2}$. On rappelle que $g_t = \frac{G_t}{n}$ est la fraction de descendant de I dans la population. On sait que $\mathbb{E}(g_{t+1}|g_t) = g_t(2-g_t)$. L'idée directrice est que si $g_t \leq \frac{1}{2}$, alors d'une génération à la suivante on peut espérer multiplier le nombre de descendants par un facteur $2 - g_t \geq \frac{3}{2}$. En particulier avec une grande probabilité, la proportion de descendants de I sera multiplié par au moins $\sqrt{2} < \frac{3}{2}$. Pour passer de n^β à $\frac{n}{2}$, seront nécessaires au maximum
$\log_{\sqrt{2}}(\frac{n^{1-\beta}}{2}) = 2[(1-\beta)\log(n) - 1]$ générations.

Proposition 5 *On suppose* $G_0 \geq n^\beta$ *et on définit* $T_3 = \inf\{t : G_t \geq \frac{n}{2}\}$. *Alors* $\mathbb{P}(T_3 > 2(1-\beta)\log(n)) = o(\frac{1}{n})$ *lorsque* $n \to \infty$.

Preuve

- Cette preuve est de structure similaire à celle de la proposition précédente. Si $n^\beta \leq G_t \leq \frac{n}{2}$ alors par l'inégalité de Bernstein on a :
  $\mathbb{P}(G_{t+1} < \sqrt{2}G_t | G_t) \leq \exp(-0.001 G_t) \leq \exp(-0.001 n^\beta))$.

- On remarque aussi que si $\log_{\sqrt{2}}(\frac{\frac{n}{2}}{n^\beta}) = 2(1-\beta)\log(n)-2$ donc si $\{T_3 > 2(1-\beta)\log(n)\}$ est réalisé, alors il existe $t < 2(1-\beta)\log(n)$ tel que $G_{t+1} \le \sqrt{2}G_t$.

- On a donc $\mathbb{P}(\{T_3 > 2(1-\beta)\log(n)\}) \le 2(1-\beta)\log(n)\exp(-0.001n^\beta)$ et $\mathbb{P}(\{T_3 > 2(1-\beta)\log(n)\}) = o(\frac{1}{n})$. $\diamond$

3.4 Étape 4

Notons $B_t = 1 - g_t$ la fraction d'individus qui non descendants de l'individu I à la génération t.Les parents d'un individu qui n'est pas descendant de I à la génération $t+1$ ne le sont pas non plus à la génération t.
Le modèle de reproduction proposé (c'est à dire deux choix uniformes et indépendants de parents) donne la relation: $(B_{t+1}|B_1,...,B_t) \sim \frac{1}{n}\ \mathrm{Bin}(n, B_t^2)$.
L'objectif de cette quatrième étape est d'estimer le temps nécessaire pour ramener B_t d'une valeur proche de $\frac{1}{2}$ à une valeur inférieure à $n^{-\alpha}$ avec $\frac{1}{2} < \frac{3\alpha}{4}$. L'idée majeure est d'utiliser que $\mathbb{E}(B_{t+1}|B_t) = B_t^2$. On peut donc s'attendre légitimement à ce que B_t soit élevé au carré d'une génération à la suivante. Plus précisément on va montrer $\lim_{t\to\infty} \mathbb{P}(B_{t+1} \ge B_t^{\frac{3}{2}}) = 0$. Il en découlera que cette quatrième étape dure au maximum $\log(\log(n))$ générations.

Proposition 6 *Soit* $\frac{1}{2} < \frac{3\alpha}{4}$ *et* $(B_t)_{t\ge 0}$ *un processus qui suit la relation de récurrence* $(B_{t+1}|B_1,...,B_t) \sim \frac{1}{n}\mathrm{Bin}(n, B_t^2)$. *On définit* $T_4 = \inf\left\{t : B_t \le \frac{1}{n^\alpha}\right\}$.
Alors $\mathbb{P}(\{T_4 \ge 2\log(\log(n))\}) = o(\frac{1}{n})$ *lorsque* $n \to \infty$.

Preuve

- D'après l'inégalité de Bernstein on a :
$$\begin{aligned}\mathbb{P}(B_{t+1} \ge B_t^{\frac{3}{2}}|B_t) &\le \exp\Big(\frac{-n^2B_t^3(1-B_t^{\frac{1}{2}})^2}{2nB_t^2(1-B_t)^2+\frac{2}{3}nB_t^{\frac{3}{2}}(1-B_t^{\frac{1}{2}}}\Big)\\ &= \exp\Big(\frac{-nB_t(1-B_t^{\frac{1}{2}})^2}{2(1-B_t)^2+\frac{2}{3}B_t^{\frac{-1}{2}}(1-B_t^{\frac{1}{2}}}\Big)\end{aligned}$$

- Si on a $\frac{1}{n^\alpha} \le B_t \le \frac{1}{2}$, alors $(1-B_t^{\frac{1}{2}})^2 \ge 1.5 - \sqrt{2} \ge 0.08$.

- Alors $\dfrac{-nB_t(1-B_t^{\frac{1}{2}})^2}{2(1-B_t)^2+\frac{2}{3}B_t^{\frac{-1}{2}}(1-B_t^{\frac{1}{2}}} \le \dfrac{-0.08n^{1-\alpha}}{2+\frac{2}{3}n^{\frac{\alpha}{2}}} \le -0.08n^{1-3/2\alpha}$ pour n assez grand.

- Cela implique $\mathbb{P}(B_{t+1} \ge B_t^{\frac{3}{2}}|B_t) \le \exp(-0.08n^{1-\frac{3\alpha}{2}})$.

- Si $B_0 \le \frac{1}{2}$ et $\forall t \in$: $[\![0, [2\log(\log(n))]]\!]$ on a $B_{t+1} \le B_t^{\frac{3}{2}}$ on peut déduire $B_{[2\log(\log(n))]} \le \dfrac{1}{n} \le \dfrac{1}{n^\alpha}$.

- En passant aux événements:
$\{T_4 > [2\log(\log(n))]\} \subseteq \left\{B_{[2\log(\log(n))]} > \frac{1}{n^\alpha}\right\} \subseteq \bigcup_{t \leq [2\log(\log(n))]-1} \left\{B_{t+1} > B_t^{3/2},\ \frac{1}{n^\alpha} < B_t < \frac{1}{2}\right\}$

- Il suit :

$$\begin{aligned}\mathbb{P}(\{T_4 > [2\log(\log(n))]\}) &\leq \sum_{t=0}^{[2\log(\log(n))]-1} \mathbb{P}(B_{t+1} \geq B_t^{\frac{3}{2}}, \frac{1}{n^\alpha} < B_t \leq \frac{1}{2}) \\ &\leq 2\log(\log(n))]\exp(-0.08n^{1-\frac{3\alpha}{2}}) = o(\frac{1}{n}). \quad \diamond\end{aligned}$$

3.5 Étape 5

Cette étape commence lorsque le processus $(B_t)_{t\geq 0}$ est inférieur à $\frac{1}{n^\alpha}$ et se termine lorsqu'il vaut exactement zéro. A la fin de cette étape, l'individu I choisi précédemment de la génération 0 est bien un ancêtre commun recherché.

Proposition 7 *Supposons* $B_0 \leq \frac{1}{n^\alpha}$. *Alors* $\lim_{n\to\infty} \mathbb{P}(B_1 = 0) = 1$.

<u>Preuve</u>
On rappelle que $B_1 \sim \frac{1}{n}Bin(n, B_0^2)$ et $2\alpha > 1$. Le calcul explicite pour la loi Binomiale nous donne: $\mathbb{P}(B_1 = 0) = (1 - B_0^2)^n \geq (1 - \frac{1}{n^{2\alpha}})^n \to 1$ $\diamond$

3.6 Étude de la borne supérieure

On montre à présent que pour $\varepsilon > 0$ fixé, $\lim_{n\to\infty} \mathbb{P}(\mathcal{T}_n > (1+\varepsilon)\log(n)) = 0$.

Proposition 8 *Soit* $\varepsilon > 0$ *fixé. Alors* $\lim_{n\to\infty} \mathbb{P}(\mathcal{T}_n > (1+\varepsilon)\log(n)) = 0$.

<u>Preuve</u>

- On utilise ici le fait que les processus $(B_t)_{t\geq 0}$ et $(G_t)_{t\geq 0}$ sont des chaînes de Markov et qu'il suffit de considérer la réalisation des étapes successives de la preuve en remplaçant à chaque fois l'hypothèse de chaque étape par la conclusion de l'étape précédente.

- On définit T_1 comme pour la première étape de la démonstration du théorème. On sait que T_1 est fini presque sûrement et qu'en plus si on fixe $\xi > 0$ arbitraire , $\lim_{n\to\infty} \mathbb{P}(T_1 > \xi \log(n)) = 0$.

- A la fin de la première étape on a identifié un individu I de la génération 0 qui a au moins $\log^2(n)$ descendants.

- On note G_t le nombre de descendants de I et $\tau(b) = \inf\{t : G_t \geq b\}$.

- Comme $(G_t)_{t\geq 0}$ et $(B_t)_{t\geq 0}$ sont des chaînes de Markov on a ainsi que les résultats précédents assure que:

- - $\mathbb{P}(\tau(n^\beta) - T_1 > \log(n)) = o(\frac{1}{n})$.
  - $\mathbb{P}(\tau(\frac{n}{2}) - \tau(n^\beta) > 2(1-\beta)\log(n) | \tau(n^\beta) < \infty)) = o(\frac{1}{n})$.
  - $\mathbb{P}(\tau(n - n^{1-\alpha}) - \tau(\frac{n}{2}) > \log(\log(n))) | \tau(\frac{n}{2}) < \infty) = o(\frac{1}{n})$.
  - $\mathbb{P}(\tau(n) - \tau(n - \frac{1}{n^\alpha}) > 1 | \tau(n - \frac{1}{n^\alpha}) < \infty) = o(1)$.

- Il en découle que :
  $\mathbb{P}(\mathcal{T}_n > \xi \log(n) + \log(n) + 2(1-\beta)\log(n) + 2\log(\log(n)) + 1) \leq$
  $\mathbb{P}(T_1 > \xi \log(n)) + \mathbb{P}(T_1 < \infty, \tau(n^\beta) - T_1 > \log(n))$ $+$
  $\mathbb{P}(\tau(n^\beta) < \infty), \tau(\frac{n}{2}) - \tau(n^\beta) > 2(1-\beta)\log(n))$ $+$
  $\mathbb{P}(\tau(\frac{n}{2}) < \infty, \tau(n - n^{1-\alpha}) - \tau(\frac{n}{2} > \log(\log(n))))$ $+$
  $\mathbb{P}(\tau(n - \frac{1}{n^\alpha}) < \infty, \tau(n) - \tau(n - \frac{1}{n^\alpha}) > 1)$.

- Donc $\mathbb{P}(\mathcal{T}_n > \xi \log(n) + \log(n) + 2(1-\beta)\log(n) + 2\log(\log(n)) + 1) = o(1)$

- Si on choisit ξ et β tel que $\xi + 2(1-\beta) < \varepsilon$ on a alors
  $\mathbb{P}(\mathcal{T}_n > (1+\varepsilon)\log(n)) \to 0$ pour $n \to \infty$

3.7 Étude de la borne inférieure

On montre dans cette partie que si on fixe $\varepsilon > 0$ alors lorsque $n \to \infty$, $\mathbb{P}(\mathcal{T}_n < (1-\varepsilon)\log(n)) \to 0$. Cela permettra de conclure par la définition de la convergence en probabilités.

Lemme 5 *Soit $\delta \leq \frac{3}{2}$ et $t \in \mathbb{N}$ un entier vérifiant l'inégalité $G_t \leq \frac{\delta n}{20}$.*
Alors $\mathbb{P}(G_{t+1} \geq (2+\delta)G_t | G_t) \leq \exp(-\frac{\delta^2 G_t}{5})$.

L'asymptotique $\mathbb{P}(\mathcal{T}_n < (1-\varepsilon)\log(n)) \to 0$ annoncée découle de ce lemme.

Proposition 9 *Soit $\varepsilon > 0$ fixé. Alors* $\lim\limits_{n\to\infty} \mathbb{P}(\mathcal{T}_n < (1-\varepsilon)\log(n) = 0$.

<u>Preuve</u>

- Fixons $\varepsilon > 0$. On veut montrer que la probabilité qu'aucun des individus de la génération 0 ne devienne un ancêtre commun avant la génération $(1-\varepsilon)\log(n)$ tend vers 0.

- On définit par récurrence sur $t \in \mathbb{N}$ le processus $(G_t)_{t\geq 0}$ par $G_0 = 1$ et G_{t+1} par la loi conditionnelle $(G_{t+1}|G_t) \sim Bin(n, \frac{2G_t}{n} - \frac{G_t^2}{n^2})$.

- On regarde $(G_t)_{t\geq 0}$ comme le processus du nombre de descendants d'un individu fixé à la génération 0. Fixons $0 < r < \varepsilon$ vérifiant l'inégalité
  $2 < 2^{\frac{1-r}{1-\varepsilon}} < 3,5$.

- On note aussi $(\tilde{G}_t)_{t\geq 0}$ le processus dont les transitions sont identiques à celle de $(G_t)_{t\geq 0}$ mais qui est réfléchit en $[n^r]$, et $\tilde{G}_0 = [n^r]$, ie :

$$(\tilde{G_{t+1}}|\tilde{G}_t) \sim max\left\{[n^r], \mathrm{Bin}(n, \frac{2\tilde{G}_t}{n} - \frac{\tilde{G}_t^2}{n^2})\right\}.$$

- Si on pose $\tau_n^G = \inf\{t : G_t = n\}$ et $\tau_n^{\tilde{G}} = \inf\left\{t : \tilde{G}_t = n\right\}$. Une étude simple d'inclusion sur les éventements donne $\mathbb{P}(\tau_n^G \geq u) \geq \mathbb{P}(\tau_n^{\tilde{G}} \geq u)$.
- Comme $\tilde{G}_0 = [n^r]$, si l'événement $\left\{\tau_n^{\tilde{G}} \leq [(1-\varepsilon)\log(n)]\right\}$ est réalisé, on peut trouver $t < [(1-\varepsilon)\log(n)]$ entier vérifiant $\tilde{G}_{t+1} \geq \tilde{G}_t 2^{\frac{1-r}{1-\varepsilon}}$.
- D'après le lemme (5) appliqué à $0 < \delta = 2^{\frac{1-r}{1-\varepsilon}} - 2 < \frac{3}{2}$ on a : $\mathbb{P}\big(\tau_n^{\tilde{G}} \leq [(1-\varepsilon)\log(n)]\big) \leq (1-\varepsilon)\log(n)\exp(-\frac{\delta^2[n^r]}{5}) = o(\frac{1}{n})$.
- On a montré que la probabilité que l'individu 1 de la génération 0 soit un ancêtre commun avant la génération $(1-\varepsilon)\log(n)$ est négligeable devant $\frac{1}{n}$. En passant à la réunion, par sous additivité on obtient que la probabilité qu'un individu soit ancêtre commun à la génération $(1-\varepsilon)\log(n)$ tend vers 0 lorsque $n \to \infty$.

Les deux dernières étapes nous donnent clairement (en appliquant la définition) la convergence en probabilités de $\frac{\mathcal{T}_n}{\log(n)}$ vers 1.

4 Preuve du Théorème 2

4.1 Déroulement de la preuve

Comme on l'a fait précédemment, on commence par donner la structure générale de la preuve. On définit $t_n = [(\zeta - \varepsilon)\log(n)]$ et $u_n = [(\zeta + \varepsilon)\log(n)]$.
Pour $1 \leq i \leq n$ le processus $(G_t^i)_{t\geq 0}$ représente le nombre de descendants de l'individu $I_{i,0}$. L'objectif est de montrer que si on attend suffisamment longtemps, tous les processus $((G_t^i)_{t\geq 0})_{i\leq n}$ prennent leurs valeurs dans $\{0; n\}$.
L'argument principal que l'on va utiliser est le suivant : avec une probabilité qui tend vers 1 lorsque $n \to \infty$, il y a une grande proportion d'entiers i tel que $G_{t_n}^i \in [1; \log^2(n)]$ et il n'y en a aucun qui vérifie que $G_{u_n}^i \in [1; \log^2(n)]$. Ce résultat découle essentiellement d'une étude approfondie du processus de Galton-Watson associé aux générations t_n et u_n.
Pour la détermination de la borne supérieure, on regarde la situation à la génération u_n. L'argument principal est qu'avec une probabilité qui tend vers 1, toutes les descendances non éteintes à la génération u_n comptent au moins $\log^2(n)$ individus. Ce nombre de descendants (pour tout individu $\tilde{I}$ de la génération 0 vérifiant la condition énoncée) est suffisamment élevé pour prédire qu'avec une probabilité qui tend vers 1, en ajoutant $(1+\varepsilon)\log(n)$ générations supplémentaires, $\tilde{I}$ deviendra un ancêtre commun.
Il en résulte $\mathcal{U}_n \leq u_n + (1+\varepsilon)\log(n)$ avec une probabilité qui tend vers 1.
Pour la borne inférieure, le point crucial est de montrer qu'à la génération t_n, une grande partie des descendances ont un cardinal qui appartient à l'intervalle $[1, \log^2(n)]$. Il est assez peu probable que tous ces processus s'éteigne simultanément. De plus comme ces

descendances ont au plus $\log^2(n)$ membres au temps t_n, avec une probabilité qui tend vers 1, il faudra au moins $(1-\varepsilon)\log(n)$ générations supplémentaires pour devenir ancêtre commun de la population. Cela nous donnera $\mathcal{U}_n > t_n+(1-\varepsilon)\log(n)$ avec une probabilité qui tend vers 1.

4.2 Un résultat sur les processus de branchement

Lemme 6 *Soit $(Y_t)_{t\geq 0}$ un processus de Galton-Watson avec $Y_0 = 1$ et une loi reproduction de Poisson de paramètre 2. Soit γ introduit dans l'énoncé du théorème 2 et une suite d'entiers strictement positifs $(b_t)_{t\geq 0}$ vérifiant $\log(b_t) = o(t)$ lorsque $t \to \infty$. Alors* $\lim\limits_{t\to\infty} \frac{\log(\mathbb{P}(1\leq Y_t\leq b_t))}{t} = \log(\gamma)$

Ce résultat est admis et se déduit de plusieurs lemmes du chapitre 1 de [2].

4.3 Preuve de la borne supérieure

Lemme 7 *Soit $I_{i,0}$ l'individu numéro i de la génération 0 et G_t^i le nombre de ses descendants à la génération t. On définit $\tau_{0,b}^i = \inf\left\{t : G_t^i = 0 \text{ ou } G_t^i \geq b\right\}$ et $A_n = \bigcup\limits_{i=1}^{n} \left\{\tau^i_{0,\log^2(n)} > (\zeta+\varepsilon)\log(n)\right\}$. Alors $\lim\limits_{n\to\infty} \mathbb{P}(A_n) = 0$*

<u>**Preuve**</u>

- On pose $\tau_{0,b}^Y = \inf\{t : Y_t = 0 \text{ ou } Y_t \geq b\}$. Comme $\left\{\tau_{0,b}^Y > t\right\} \subseteq \{1 \leq Y_t < b\}$ le lemme 6 assure qu'en supposant $\log(b) = o(t)$ et $tb^2 = o(n)$, il vient $\limsup\limits_{t\to\infty} \frac{\log \mathbb{P}(\tau^i_{0,b}>t)}{t} \leq \log(\gamma)$.
- On fixe $\varepsilon > 0$ et on applique l'inégalité que l'on vient de montrer à la génération $t = [(\zeta+\varepsilon)\log(n)]$ et $b = [\log^2(n)]$. On a alors pour $\delta > 0$ fixé et n suffisamment grand :
$\log(\mathbb{P}(\tau^i_{0,\log^2(n)} > (\zeta+\varepsilon)\log(n))) \leq (\log(\gamma)+\delta)(\zeta+\varepsilon)\log(n)$.
- Si $\delta > 0$ est suffisamment petit, par passage à la fonction exponentielle $\mathbb{P}(\tau^i_{0,\log^2(n)} > (\zeta+\varepsilon)\log(n)) = o(\frac{1}{n})$. En utilisant la sous additivité on a $\mathbb{P}(A_n) = o(1)$ $\diamond$

On vient de démontrer qu'avec une probabilité qui tend vers 1, les individus de la génération 0 ont tous soit 0 descendants soit au moins $\log^2(n)$ descendants à la génération $[(\zeta+\varepsilon)\log(n)]$.
A présent nous allons montrer que pour $\varepsilon > 0$ fixé, chaque individu qui a au moins $\log^2(n)$ descendants à la génération $[(\zeta+\varepsilon)\log(n)]$ deviendra un ancêtre commun après $(1+\varepsilon)\log(n)$ générations supplémentaires.
La majorité du travail a déjà été réalisé pendant la preuve du premier théorème. Il faudra cependant ajouter un raffinement supplémentaire à l'étape 5 de la preuve du premier théorème. On rappelle que $B_t = 1 - g_t$.

Lemme 8 *Soit $\alpha \in]\frac{1}{2}, \frac{2}{3}[$ fixé et on se donne un entier $k(\alpha) > \frac{1}{2\alpha-1}$.*
On suppose que $B_0 \leq n^{-\alpha}$ et on définit $T_5 = \inf\{t : B_t = 0\}$.
Alors $\mathbb{P}(T_5 > k(\alpha)) = o(\frac{1}{n})$ lorsque $n \to \infty$.

Preuve

- On rappelle que le processus $(B_t)_{t\geq 0}$ est définit par $(B_{t+1}|B_t) \sim \frac{1}{n}\mathrm{Bin}(n, B_t^2)$.
  Si $B_t \leq \frac{1}{n^\alpha}$ et que n est suffisamment grand :
  $$\begin{aligned}\mathbb{P}(B_{t+1} > 0|B_t) &= 1 - (1 - B_t^2)^n \\ &\leq 1 - (1 - 2nB_t) \leq 2n^{1-2\alpha}\end{aligned}$$

- Lorsque $n \to \infty$ on a $\mathbb{P}(0 < B_{t+1} \leq \frac{1}{n^\alpha} | 0 < B_t \leq \frac{1}{n^\alpha}) \leq 2n^{1-2\alpha}$.

- Ensuite par l'inégalité de Bernstein appliquée à l'événement $\{B_t \leq \frac{1}{n^\alpha}\}$, on a :
  $$\begin{aligned}\mathbb{P}(B_{t+1} > 0|B_t) &= \mathbb{P}(\tfrac{\mathrm{Bin}(n,B_t^2)}{n} > B_t^2 + (n^{-\alpha} - B_t^2)|B_t) \\ &\leq \exp\big(\tfrac{-n^2(n^{-\alpha}-B_t^2)^2}{2nB_t^2(1-B_t^2)+\frac{2}{3}n(n^{-\alpha}-B_t^2}\big) \\ &\leq \exp\tfrac{-n^{2-2\alpha}}{2n^{1-2\alpha}+\frac{2}{3}n^{-\alpha}}\end{aligned}$$

- Pour n suffisamment grand et l'événement $\{B_t \leq n^{-\alpha}\}$ réalisé,
  $\mathbb{P}(B_{t+1} > 0|B_t) \leq \exp(-n^{1-\alpha})$.

- A présent on suppose que $\{B_0 \leq n^{-\alpha}\}$ est réalise. Il vient alors que
  $\bigcup\limits_{t=0}^{k} \{B_t > n^{-\alpha}\} \subseteq \bigcup\limits_{t=0}^{k-1} \{B_t \leq n^{-\alpha}, B_{t+1} > n^{-\alpha}\}$.
  Comme $\mathbb{P}(B_t \leq n^{-\alpha} \cap B_{t+1} > n^{-\alpha}) = \mathbb{P}(B_t \leq n^{-\alpha})\mathbb{P}(B_{t+1} > n^{-\alpha}|B_t)$ on déduit facilement $\mathbb{P}(B_t \leq n^{-\alpha} \cap B_{t+1} > n^{-\alpha}) \leq \exp(-n^{1-\alpha})$. Puis la sous-additivité donne $\mathbb{P}(\bigcup\limits_{t=0}^{k-1} \{B_t > n^{-\alpha}, B_{t+1} > n^{-\alpha}\}) \leq k\exp(-n^{1-\alpha})$.

- En recombinant les équations précédemment obtenues on a :
  $$\begin{aligned}\mathbb{P}(T_5 > k) &\leq \mathbb{P}(\bigcup_{t=0}^{k} \{B_t > n^{-\alpha}\}) + \mathbb{P}(\bigcap_{t=0}^{k} \{0 < B_t < n^{-\alpha}\}) \\ &\leq k\exp(-n^{1-\alpha}) + (2n^{1-\alpha})^k\end{aligned}$$

- Pour $k > \frac{1}{2\alpha-1}$ on a la majoration désirée.

On va maintenant démontrer la borne inférieure du résultat.

Preuve de la borne inférieure

- On rappelle que $\mathcal{U}_n$ est la génération à laquelle chacun des individus de la génération 0 est devenu soit un AC soit n'a plus de descendance. On note A_n comme précédemment, on définit $\tau^i(b) = \inf\{t : G_t^i \geq b\}$.

- $$\begin{aligned}\{\mathcal{U}_n > (1+\zeta+2\varepsilon)\log(n))\} &\subseteq A_n \cup [A_n^c \bigcup \{\mathcal{U}_n > (1+\zeta+2\varepsilon)\log(n)\}] \\ &\subseteq A_n \cup \bigcup_{i=1}^{n} \Big\{\tfrac{\tau^i(\log^2(n))}{\log(n)} \leq (\zeta+\varepsilon) \bigcap \tfrac{\tau^i(n)}{\log(n)} > (1+\zeta+2\varepsilon)\Big\}\end{aligned}$$

- Pour démontrer le résultat il suffit de montrer que pour $1 \leq i \leq n$:
$\mathbb{P}(\tau^i(\log^2(n)) \leq (\zeta+\varepsilon)\log(n) \cap \tau^i(n) > (1+\zeta+2\varepsilon)\log(n)) = o(\frac{1}{n})$

- Si on applique à nouveau les résultats obtenus dans la première preuve on obtient :

  - $\mathbb{P}(\tau^i(n^\beta) - \tau^i(\log^2(n)) > \log(n)|\tau^i(\log^2(n))) = o(\frac{1}{n})$
  - $\mathbb{P}(\tau^i(\frac{n}{2}) - \tau^i(n^\beta) > 2(1-\beta)\log(n)|\tau^i(n^\beta) < \infty)) = o(\frac{1}{n})$
  - $\mathbb{P}(\tau^i(n-n^{1-\alpha}) - \tau^i(\frac{n}{2}) > 2\log(\log(n)))|\tau^i(\frac{n}{2}) < \infty) = o(\frac{1}{n})$

- Le lemme 8 assure $\mathbb{P}(\tau^i(n) - \tau^i(n-\frac{1}{n^\alpha}) > k(\alpha)|\tau^i(n-\frac{1}{n^\alpha}) < \infty) = o(\frac{1}{n})$.

- Donc $\mathbb{P}(\tau^i(n) - \tau^1(n-n^{1-\alpha}) > S_{n,\alpha,\beta}|\tau^i(n-n^{1-\alpha}) < \infty) = o(\frac{1}{n})$
avec $S_{n,\alpha,\beta} = \log(n) + 2(1-\beta)\log(n) + 2\log(\log(n)) + k(\alpha)$.

- Si on choisit β suffisamment proche de 1, pour $\varepsilon > 0$ assez petit,
$\mathbb{P}(\tau^i(n) - \tau^i(\log^2(n)) > (1+\varepsilon)\log(n)|\tau^i(\log^2(n)) < \infty) = o(\frac{1}{n})$.

- Finalement $\mathbb{P}(\frac{\tau^i(\log^2(n))}{\log(n)} \leq (\zeta+\varepsilon) \cap \frac{\tau^i(n)}{\log(n)} > (1+\zeta+2\varepsilon))$
$\leq \mathbb{P}(\tau^i(\log^2(n)) < \infty, \tau^i(n) - \tau^i(\log^2(n)) > (1+\varepsilon)\log(n)) = o(\frac{1}{n})$.

- On conclut en utilisant la sous additivité.

4.4 Preuve de la borne inférieure

On va étudier dans cette partie le temps minimal nécessaire pour réaliser la condition du second théorème. On commence par montrer qu'à la génération $t_n = [(\zeta-\varepsilon)\log(n)]$, il y a plusieurs individus de la population initiale qui vérifient $1 \leq G^i_{t_n:} \leq \log^2(n)$. La probabilité que leur descendance s'éteigne pour ces individus tend vers 0 lorsque $n \to \infty$. Dans le cas (le plus probable) où toutes les descendances associées ne s'éteignent pas, il faut attendre au moins $(1-\varepsilon)\log(n)$ générations pour espérer trouver un ancêtre commun.
On rappelle que $t_n = [(\zeta-\varepsilon)\log(n)]$ et on prend les notations suivantes :
Pour $1 \leq i \leq n$, on note $J_i = \left\{G^i_t \in [1, \log^2(n)] \ \forall t \leq t_n\right\}$. Par abus de notation on appelle J_i l'indicatrice de J_i. Autrement dit $J_i = 1$ est l'événement: la descendance de i n'est pas éteinte à t_n mais qu'elle ne compte pas plus de $\log^2(n)$ individus. Au temps t_n les individus qui vérifient $J_i = 1$ ont toujours une chance de devenir un ancêtre commun, mais d'après le théorème 1, cela prendra au moins $\log(n)$ générations .
On note pour la suite $N_n = \sum\limits_{i=1}^{n} J_i$. Le lemme suivant montre qu'il n'y a que peu de dépendance (pour les premières générations) entre la taille des descendances de deux individus différents de la génération zéro. On se contente de fournir un résultat de majoration de la corrélation.

Lemme 9 $\mathbb{P}(J_1 \cap J_2) \leq (\mathbb{P}(J_1))^2(1+o(1))$ *lorsque* $n \to \infty$.

<u>**Preuve**</u>

- On considère $I_{1,0}$ et $I_{2,0}$ les individus 1 et 2 de la génération 0. On note A_t le nombre de descendants de $I_{1,0}$ mais pas de $I_{2,0}$ à la génération t. De même on note C_t le nombre de descendants de $I_{2,0}$ mais pas de $I_{1,0}$ à la génération t. Enfin on note B_t le nombre de descendants communs à $I_{1,0}$ et $I_{2,0}$ de à la génération t de (ici B_t n'a pas de lien avec celui introduit dans la preuve du premier théorème).

- Il est clair que $A_t + B_t = G_t^1$ et $C_t + B_t = G_t^2$. Le modèle de reproduction choisit assure que $(H_t)_{t\geq 0} = \big((A_t, B_t, C_t)\big)_{t\geq 0}$ est une chaîne de Markov.

- Pour simplifier les notations, on note $\mathbb{P}_H(a_t, b_t, c_t) = \mathbb{P}(H_t = (a_t, b_t, c_t))$ et $\mathbb{P}_H(a_{t+1}, b_{t+1}, c_{t+1}|a_t, b_t, c_t)$ $\mathbb{P}(H_{t+1} = (a_{t+1}, b_{t+1}, c_{t+1})|H_t = (a_t, b_t, c_t))$.

- On va commencer par démontrer que $\mathbb{P}(J_1 \cap J_2) \sim \mathbb{P}(J_1 \cap J_2 \cap \{B_t = 0\ \forall t \leq t_n\})$ lorsque $n \to \infty$.

- Intuitivement, si A_t et C_t sont majorés par $\log^2(n)$ et que $B_t = 0$, alors la probabilité conditionnelle de $\{B_{t+1} > 0\}$ est majorée par $\frac{2A_tC_t}{n}\mathcal{O}(\frac{\log^4(n)}{n})$.

- Cela suggère que si $J_1 \cap J_2$ est réalisé, alors pour $s \leq t_n$, la probabilité conditionnelle que $\{B_t > 0\}$ pour la première fois à la génération s est $\mathcal{O}(\frac{\log^4}{n})$. Si on ajoute ces probabilités pour les $s \leq t_n = \mathcal{O}(\log(n))$ cela donne $\mathbb{P}(\exists t \leq t_n, B_t > 0|J_1 \cap J_2) = \mathcal{O}(\frac{\log^5(n)}{n})$.

- Cela implique $\mathbb{P}(J_1 \cap J_2) = \mathbb{P}(J_1 \cap J_2 \cap \{B_t = 0\ \forall t \leq t_n\})\big[1 + \mathcal{O}(\frac{\log^5(n)}{n})\big]$ et donne le résultat attendu.

- Passons à la preuve formelle. On note $\mathcal{L}_n = [1; \log^2(n)]$. On a alors :
$$\begin{aligned}\mathbb{P}(J_1 \cap J_2|\,\{B_t = 0 \forall t \leq t_n\}) &= \sum_{a_i\in\mathcal{L}_n, c_j\in\mathcal{L}_n, i\leq t_n, j\leq t_n} \mathbb{P}_H(a_1, 0, c_1, a_2, 0, c_2,\ \dots\ , a_n, 0, c_n)\\ &= \sum_{a_1,c_1\in\mathcal{L}_n} \mathbb{P}_H(a_1, 0, c_1) \sum_{a_2,c_2\in\mathcal{L}_n} \mathbb{P}_H(a_2, 0, c_2|a_1, 0, c_1)\ \dots\\ &\quad \sum_{a_{t_n},c_{t_n}\in\mathcal{L}_n} \mathbb{P}_H(a_{t_n}, 0, c_{t_n}|a_{t_{n-1}}, 0, c_{t_{n-1}})\end{aligned}$$

- On définit $\alpha_s = \dfrac{2a_s}{n} - \dfrac{a_s(a_s + 2c_s)}{n^2}$, $\beta_s = \dfrac{2a_sc_s}{n^2}$ et $\gamma_s = \dfrac{2c_s}{n} - \dfrac{c_s(c_s + 2a_s)}{n^2}$.

- Le conditionnement par les valeurs à la génération $t-1$ donne :
$$\begin{aligned}\mathbb{P}_H(a_t, 0, c_t|a_{t-1}, 0, c_{t-1}) &= \mathbb{P}(\mathrm{Bin}(n, \alpha_{t-1}) = a_t)\mathbb{P}(\mathrm{Bin}(n - \alpha_t, \tfrac{\gamma_{t-1}}{1-\alpha_{t-1}}) = c_t)\\ &\quad \mathbb{P}(\mathrm{Bin}(n - a_t - c_t, \tfrac{\beta_{t-1}}{1-\alpha_{t-1}-\gamma_{t-1}}) = 0)\end{aligned}$$

- On veut comparer ceci à la probabilité analogue pour deux processus indépendants $(G_t)_{t\geq 0}$ de même loi de transition, i.e qui vérifient :
$\mathbb{P}_G(a_t|a_{t-1})\mathbb{P}_G(c_t|c_{t-1}) = \mathbb{P}(\mathrm{Bin}(n, \alpha_{t-1} + \beta_{t-1} = a_t)\mathbb{P}(\mathrm{Bin}(n, \gamma_{t-1} + \beta_{t-1} = c_t)$.

- Le quotient $\dfrac{\mathbb{P}_H(a_t,0,c_t|a_{t-1},0,c_{t-1})}{\mathbb{P}_G(a_t|a_{t-1})\mathbb{P}_G(c_t|c_{t-1})}$ est le produit des termes suivants: $\dfrac{\mathbb{P}(\mathrm{Bin}(n,\alpha_{t-1})=a_t)}{\mathbb{P}(\mathrm{Bin}(n,\alpha_{t-1}+\beta_{t-1})=a_t)}$, $\dfrac{\mathbb{P}(\mathrm{Bin}(n-a_t,\frac{\gamma_{t-1}}{1-\alpha_{t-1}})=c_t)}{\mathbb{P}(\mathrm{Bin}(n,\gamma_{t-1}+\beta_{t-1})=c_t)}$ et $\mathbb{P}(\mathrm{Bin}(n-a_t-c_t,\dfrac{\beta_{t-1}}{1-\alpha_{t-1}-\gamma_{t-1}})=0)$

- On majore le dernier terme par 1 (il est proche de 1 donc on ne perd pas beaucoup d'information).

- Le premier terme vaut après simplifications $\dfrac{\alpha_{t-1}^{a_t}(1-\alpha_{t-1})^{n-a_t}}{(\beta_{t-1}+\alpha_{t-1})^{a_t}(1-\alpha_{t-1}-\beta_{t-1})^{n-a_t}} \leq \big(1+\dfrac{\beta_{t-1}}{1-\alpha_{t-1}-\beta_{t-1}}\big)^n = 1+\mathcal{O}(\dfrac{\log^4(n)}{n})$ car $\dfrac{\beta_{t-1}}{1-\alpha_{t-1}-\beta_{t-1}} \sim \beta_{t-1} \leq \dfrac{2\log^4(n)}{n^2}$.

- Par des calculs similaires $\dfrac{\mathbb{P}(\mathrm{Bin}(n-a_t,\frac{\gamma_{t-1}}{1-\alpha_{t-1}})=c_t)}{\mathbb{P}(\mathrm{Bin}(n,\gamma_{t-1}+\beta_{t-1})=c_t)} = 1+\mathcal{O}(\dfrac{\log^4(n)}{n})$.

- En multipliant les majorations $\dfrac{\mathbb{P}_H(a_t,0,c_t|a_{t-1},0,c_{t-1})}{\mathbb{P}_G(a_t|a_{t-1})\mathbb{P}_G(c_t|c_{t-1})} = 1+\mathcal{O}(\dfrac{\log^4(n)}{n})$.

- On a donc :

$$\begin{aligned}
\mathbb{P}(J_1\cap J_2|\{B_t=0\forall t\leq t_n\}) &= \sum_{a_i\in\mathcal{L}_n,c_j\in\mathcal{L}_n,i\leq t_n,j\leq t_n} \mathbb{P}_H(a_1,0,c_1,a_2,0,c_2,\ \dots\ ,a_n,0,c_n)\\
&= \sum_{a_1,c_1\in\mathcal{L}_n}\mathbb{P}_H(a_1,0,c_1)\sum_{a_2,c_2\in\mathcal{L}_n}\mathbb{P}_H(a_2,0,c_2|a_1,0,c_1)\ \dots\\
&\quad \sum_{a_{t_n},c_{t_n}\in\mathcal{L}_n}\mathbb{P}_H(a_{t_n},0,c_{t_n}|a_{t_{n-1}},0,c_{t_{n-1}})\\
&\leq \sum_{a_1,c_1\in\mathcal{L}_n}\mathbb{P}_G(a_1|1)\mathbb{P}_G(c_1|1)\sum_{a_2,c_2\in\mathcal{L}_n}\mathbb{P}_G(a_2|a_1)\mathbb{P}_G(c_2|c_1)\\
&\quad \sum_{a_{t_n},c_{t_n}\in\mathcal{L}_n}\mathbb{P}_G(a_n|a_{n-1})\mathbb{P}_G(c_n|c_{n-1})[1+\mathcal{O}(\tfrac{\log^4(n)}{n})]^{t_n}\\
&= [\mathbb{P}(G_1\in\mathcal{L}_n\quad\forall t\geq t_n)]^2[1+\mathcal{O}(\tfrac{\log^4(n)}{n})]^{t_n}\\
&= \mathbb{P}(J_1)^2[1+\mathcal{O}(\tfrac{\log^5(n)}{n})]
\end{aligned}$$

- Ceci termine la preuve $\diamond$

On montre maintenant le dernier lemme nécessaire à la preuve de la borne inférieure

Lemme 10 *Lorsque $n\to\infty$, alors $N_n\to\infty$ en probabilités.*

<u>Preuve</u>

- On va montrer que la moyenne et de la déviation standard vérifient $\mathbb{E}(N_n)\to\infty$ et $SD(N_n)=o(\mathbb{E}(N_n))$. Par linéarité $\mathbb{E}(N_n)=n\mathbb{P}(J_1)$ (on rappelle que $\mathbb{P}(J_1)=\mathbb{P}(Y_t\in[1,\log^2(n)]\ \forall t\leq t_n)$).

- On remarque que $\mathbb{P}(Y_t \in [1, \log^2(n)]\ \forall t \leq t_n)$ peut s'approcher par $\mathbb{P}(Y_{t_n} \in [1, \log^2(n)])$ lorsque n tend vers $+\infty$. En effet si on pose $\Delta_n = \mathbb{P}(Y_{t_n} \in [1, \log^2(n)]) - \mathbb{P}(Y_t \in [1, \log^2(n)]\ \forall t \leq t_n)$ on remarque $\Delta_n = \mathbb{P}(Y_{t_n} \in [1, \log^2(n)] \cap \exists t \leq t_n\ Y_t > \log^2(n))$.

- Cette probabilité correspond à la probabilité de dépasser $\log^2(n)$ avant la génération t_n puis redescendre en dessous du niveau $\log^2(n)$ à la génération t_n. L'inégalité de Bernstein appliqué à la distribution de Poisson donne $\mathbb{P}(Y_{t+1} \leq Y_t | Y_t) \leq \exp(-\frac{3}{14} Y_t) \leq \exp(-\frac{3}{14}\log^2(n))$ si l'événement $\left\{Y_t > \log^2(n)\right\}$ est réalisé.

- Par sous additivité, $|\Delta_n| \leq t_n \exp(-\frac{3}{14}\log^2(n)) = o(\frac{1}{n})$.

- D'après le lemme 6, $\dfrac{\log(\mathbb{P}(Y_{t_n} \in [1; \log^2(n)]))}{(\zeta - \epsilon)\log(n)} \sim \dfrac{\mathbb{P}(Y_{t_n} \in [1; \log^2(n)])}{t_n}$ et $\lim\limits_{n\to\infty} \dfrac{\mathbb{P}(Y_{t_n} \in [1; \log^2(n)])}{t_n} = \mathrm{l}$
En passant à l'exponentielle on retrouve $\lim\limits_{n\to\infty} n\mathbb{P}(Y_{t_n} \in [1; \log^2(n)]) = +\infty$.

- Finalement $n\mathbb{P}(J_1) = n[\mathbb{P}(Y_{t_n} \in [1; \log^2(n)]) + o(\frac{1}{n})] \to +\infty$.

- Pour terminer la démonstration, on doit montrer que la déviation est négligeable devant $\mathbb{E}(N_n)$. Or par définition on a :
$$\begin{aligned}\mathrm{Var}(N_n) = \mathbb{E}(N_n^2) - \mathbb{E}(N_n)^2 &= n\mathbb{P}(J_1) + n(n-1)\mathbb{P}(J_1 J_2) - (n\mathbb{P}(J_1))^2 \\ &\leq n\mathbb{P}(J_1) + n(n-1)\mathbb{P}(J_1)^2(1 + o(1)) - (n\mathbb{P}(J_1))^2 \\ &= o(n^2\mathbb{P}(J_1)^2) = o(\mathbb{E}(N_n)^2)\end{aligned}$$

- Ceci conclut la preuve. $\diamond$

<u>Preuve de la borne inférieure</u>

- On définit $W_n = \left\{i : G^i_{t_n} \in [1; \log^2(n)]\right\}$. Alors on peut écrire :
$$\begin{aligned}\mathbb{P}(\mathcal{U}_n \leq t_n + (1-\varepsilon)\log(n) &\leq \mathbb{P}(W_n = \emptyset) + \mathbb{P}(\mathrm{G_i}\ \mathrm{s'eteint}\ \forall i \in W_n) \\ &+ \mathbb{P}(\exists i \in W_n\ ,\ G^i_{[t_n + (1-\epsilon)\log(n)]} = n)\end{aligned}$$

- Le cardinal de W_n est $N_n \to \infty$ en probabilité. Des résultats d'un article annexe montrent facilement que la probabilité que la descendance de tous les individus de W_n s'éteigne tend vers 0.

- Reste à montrer que $\lim\limits_{n\to\infty} \mathbb{P}(\exists i \in W_n\ ,\ G^i_{[t_n+(1-\epsilon)\log(n)]} = n) = 0$.

- Pour $i \in W_n$, on observe que si l'événement $\left\{G^i_{[t_n+(1-\epsilon)\log(n)]} = n\right\}$ est réalisé, le processus $(G^i_t)_{t\geq 0}$ doit être inférieur à $\log^2(n)$ à la génération t_n et vaut n a la génération $t_n + (1-\epsilon)\log(n)]$.

- Cela implique que le processus $(G^i_t)t \geq 0$ doit croître de $\log^2(n)$ à n en moins de $(1-\epsilon)\log(n)$ générations.

- D'après la preuve du premier théorème, ceci arrive avec une probabilité qui est $o(\frac{1}{n})$.

- On conclut par sous-additivité. ◇

La démonstration de la borne inférieure et de la borne supérieure donne la convergence en probabilités simplement en appliquant la définition.

5 Simulations

On va maintenant simuler le processus ainsi que plusieurs variantes, pour vérifier la cohérence des résultats théoriques avec la pratique.

5.1 Simulation du Théorème 1

On simule l'évolution de la généalogie grâce au code Python **Reproduction1** disponible en annexe. On moyenne la valeur de $\mathcal{T}_n$ sur 20 lancers. La courbe est un lissage sur l'ensemble des multiples de 1000 entre 1000 et 50000.

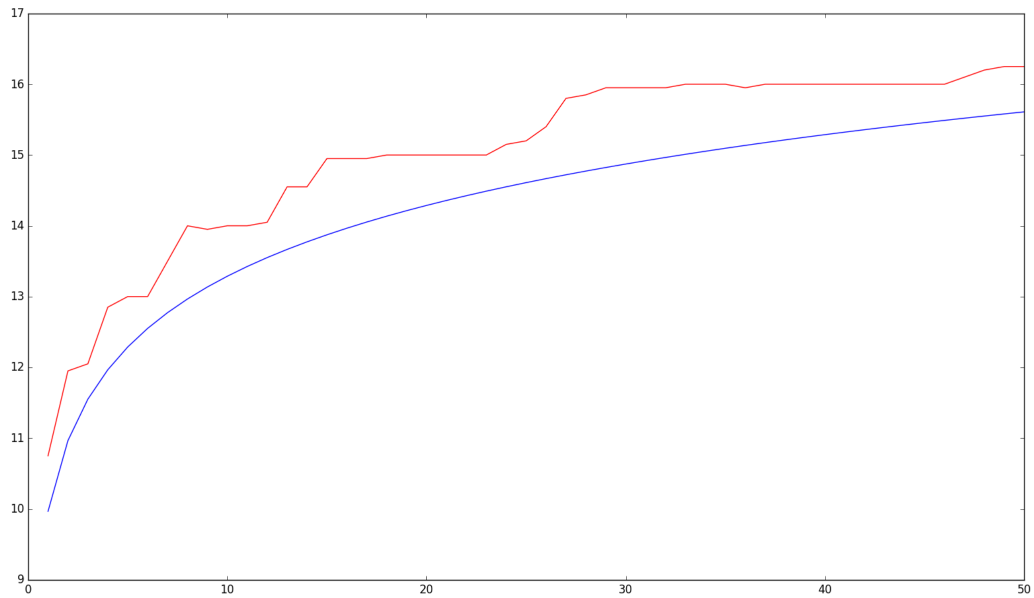

Figure 1: $\mathcal{T}_n$ en fonction de $1000n$. En rouge la courbe empirique, en bleu le logarithme en base 2

5.2 Simulation du Théorème 2

On simule l'évolution de la généalogie grâce au code Python **Reproduction2** disponible en annexe. On moyenne la valeur de $\mathcal{U}_n$ sur 20 lancers. La courbe est un lissage sur l'ensemble des multiples de 1000 entre 1000 et 20000.

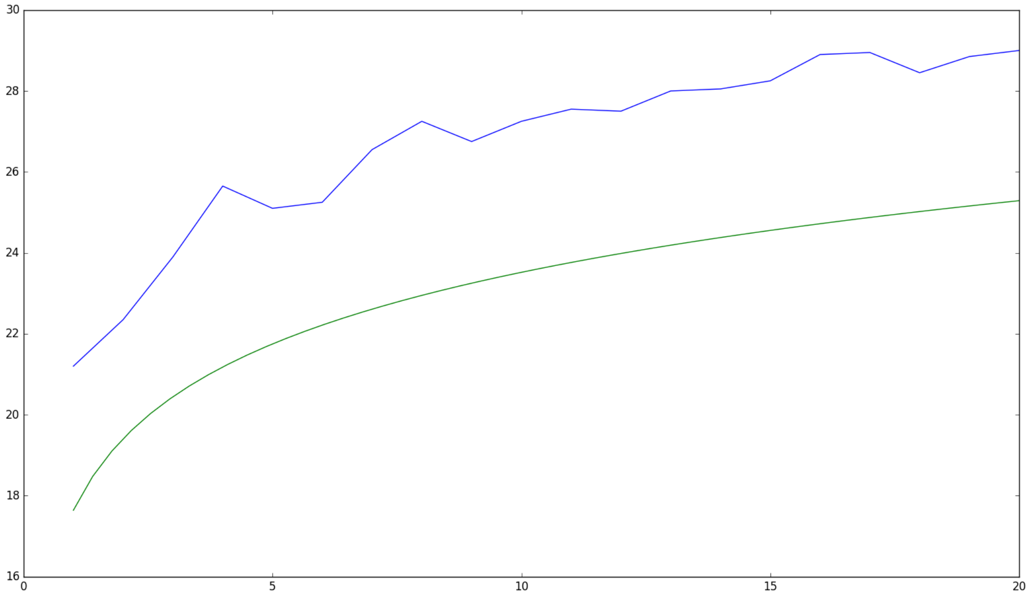

Figure 2: $\mathcal{U}_n$ en fonction de $1000n$. En rouge la courbe empirique, en bleu $1.77\log_2(x)$

5.3 Simulation de la prise en compte du sexe des parents

On simule l'évolution de la généalogie en prenant en compte le sexe des parents grâce au code Python **Reproduction3** disponible en annexe. On se donne $0 < \alpha < 1$. A chaque génération, la population est constitué d'une fraction α de femelles et $1-\alpha$ de mâles. Le choix des parents se fait alors se fait alors de manière équiprobable et indépendants. Les fractions de mâles et femelles restent constantes. On cherche à nouveau le nombre de générations nécessaires pour trouver un PRAC. On remarque que l'on obtient (d'après les simulations) toujours un PRAC et que ce PRAC arrive plus rapidement qu'avec le modèle présenté dans la première partie.
Dans le graphique qui suit, on observe l'influence de a sur $\mathcal{T}_n$. Le cas ou il faut attendre le plus de générations semble être $a = 0.5$.

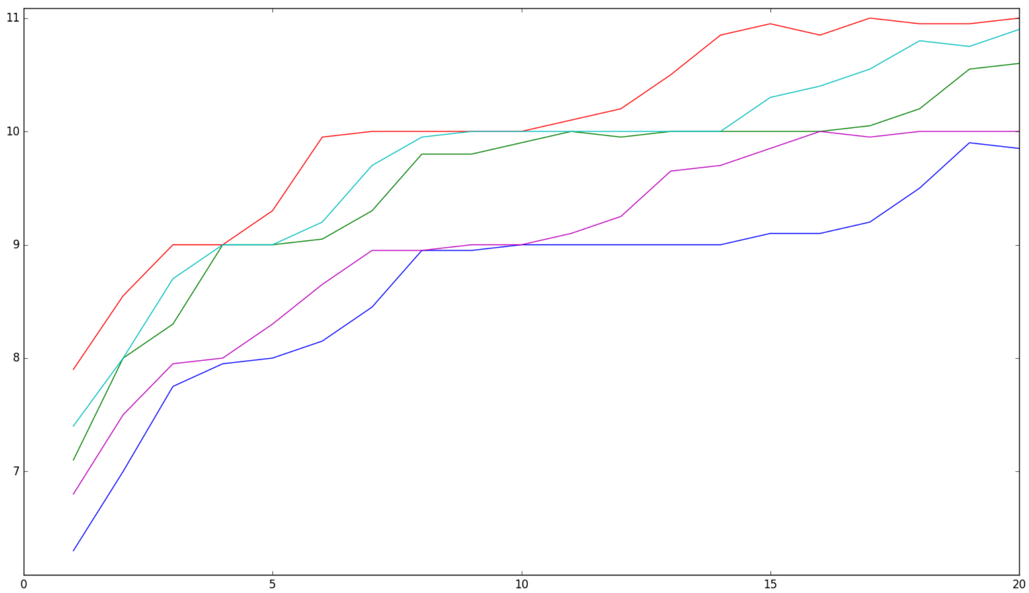

Figure 3: $\mathcal{T}_n$ en fonction de $1000n$. En bleu $a = 0.1$, en violet $a = 0.9$, en rouge $a = 0.5$n, en vert $a = 0.25$, en bleu clair $a = 0.75$

6 Annexe

```
def Reproduction1 (n):
    L=[1]*n
    T=0
    Termine = False
    A=L
    B=L
    for k in range (0,2*n):
        T=k
        if (Termine == False):

            for i in range(0,n):
                B[i]=rd.binomial(n, 2*A[i]/n - A[i]*A[i]/(n*n))

            m=max(B)
            A=B
            if (m==n):
                Termine =True
```

```
        else:
            return(T,B);

def Reproduction2 (n):
    L=[1]*n
    T=0
    Termine = False
    A=L
    B=L
    for k in range (0,2*n):
        T=k
        s=0
        if (Termine == False):

            for i in range(0,n):
                B[i]=rd.binomial(n, 2*A[i]/n - A[i]*A[i]/(n*n))
                if (B[i]==0 or B[i]==n):
                    s+=1

            A=B
            if (s==n):
                Termine =True


        else:
            return(T,B);

def Reproduction3(n,a):
    f=math.floor(n*a)
    m=n-f
    d=0
    L=[]
    for k in range(0,f):
        L.append([1,0])
    for k in range(0,m):
        L.append([0,1])
    T=0
    Termine = False
    A=L
    B=L
    C=[0]*(n)
```

```
for k in range (0,n):
    T=k
    if (Termine == False):
        for i in range(0,n):
            B[i][0]=rd.binomial(f, (A[i][0])/f+(A[i][1])/m - (A[i][0]*A[i][1])/(f*m))
            B[i][1]=rd.binomial(m, (A[i][0])/f+(A[i][1])/m - (A[i][0]*A[i][1])/(f*m))
            C[i]=B[i][0]+B[i][1]
        d=max(C)
        A=B
        if (d==n):
            Termine =True


    else:
        return(T,B);
```

7 Bibliographie

[1] Chang, J.(1999). Recent common ancestor of all present-day individuals. Adv. Appl. Prob. 31, 1002–1026.
[2] Athreya, K. B. and Ney, P. (1972). Branching Processes. Springer, New York.
[3] Hudson, R. R. (1990). Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7,1–44.