

Mémoire de L3
Nombre de chemins croissants dans un arbre
 N -aire

Nicolas HEITZ et Côme LE REVEILLE
sous la direction de Bastien MALLEIN

2023

Résumé

A partir d'un modèle de biologie évolutive, nous étudions un arbre N -aire aléatoire. A chaque noeud σ , on associe une variable aléatoire x_σ et on considère uniquement les branches de l'arbre où les variables sont croissantes avec la profondeur. On s'intéresse alors au comportement asymptotique de ce modèle lorsque N tend vers $+\infty$, en présentant majoritairement des résultats issus de [Che14]. On obtient d'une part un ordre de grandeur pour la population à la génération αN et d'autre part, une estimée de la profondeur à laquelle l'arbre s'arrête : $eN - \frac{3}{2} \ln(N)$. Pour conclure, on propose une modélisation informatique de ces objets, afin d'observer graphiquement certains des résultats obtenus.

Table des matières

1	Motivation et description du modèle d'arbre aléatoire	2
1.1	Première modélisation du mécanisme d'évolution et de sélection des génomes	2
1.2	Le modèle de l'arbre N -aire	3
2	Premiers calculs concernant le nombre de sommets accessibles	4
2.1	Espérance du nombre de sommets accessibles	4
2.2	Fonction génératrice du nombre de sommets accessibles	5
3	Résultats principaux sur l'asymptotique du nombre de chemins croissants	6
3.1	Comportement asymptotique du nombre de chemins accessibles au premier ordre	6
3.2	Génération critique d'extinction	7
3.3	Adaptation au cas où la racine est initialisée à une valeur $x \neq 0$	7
4	Preuves	9
4.1	Lemmes	9
4.2	Théorème de convergence de $Z_{N,\alpha N}$	12
4.3	Estimation asymptotique du temps d'extinction $M_N^{(0)}$ au second ordre	14
4.4	Arbre à loi de reproduction binomiale	17
5	Discrétisation et modélisation de l'arbre N-aire	18
5.1	Discrétisation des valeurs de survivabilité	18
5.2	Algorithme de modélisation	20
5.3	Observations	21

Chapitre 1

Motivation et description du modèle d'arbre aléatoire

1.1 Première modélisation du mécanisme d'évolution et de sélection des génomes

Le but originel du modèle mathématique que nous allons présenter est de décrire de manière simplifiée les mécanismes de sélection qui ont lieu au cours de l'évolution d'une espèce. Il est dérivé d'un modèle proposé dans "*How many evolutionary histories only increase fitness?*" ([BBS16]) qui part de la description suivante : on considère un génome simplifié avec N gènes où chaque gène possède deux allèles, notés 0 et 1, le premier étant l'allèle originel et le second, l'allèle mutant. On peut donc représenter chaque génotype par un N -uplet dans $\{0; 1\}^N$. A chaque génome (donc à chaque N -uplet), on attribue une valeur aléatoire de survivabilité (*fitness*) entre 0 et 1. Celle-ci correspond à l'aptitude à la survie des individus. Les conditions sont alors les suivantes : le génôme $(1, 1, \dots, 1)$ possède une survivabilité de 1 et tous les autres génômes ont une survivabilité tirée de manière aléatoire et indépendante dans $[0; 1]$. Ce choix peut paraître intrigant, mais de nombreuses observations biologiques et modèles mathématiques privilégient de telles modélisations. En effet le modèle *House of cards* présente l'évolution comme l'effet de petites mutations qui entraînent une cascade de modifications dans l'organisme de l'individu (et ce, relativement aux autres allèles déjà présents, comme présenté par exemple par Kingman [Kin78]). Et nous choisissons donc de modéliser cet aléa des modifications par des valeurs de survivabilité uniformes, aléatoires et indépendantes entre elles.

Aux conditions précédemment présentées s'ajoute une hypothèse de forte sélection, qui impose que toute mutation défavorable ne peut subsister tandis que toute mutation favorable vient supplanter l'intégralité de l'espèce. Ainsi, ce modèle s'intéresse aux déplacements croissants (pour la survivabilité et le nombre de mutations) dans un hypercube de dimension N , partant du point $(0, 0, \dots, 0)$ jusqu'au sommet opposé $(1, 1, \dots, 1)$. En résumé, en notant $\sigma_0 = (0, 0, \dots, 0)$ et $\sigma_N = (1, 1, \dots, 1)$, on s'intéresse aux chemins σ de longueur N où σ_{i+1} s'obtient à partir de σ_i en changeant un 0 en 1, et où $x_{\sigma_0} < x_{\sigma_1} < \dots < x_{\sigma_N}$ (x_{σ_i} correspondant à la valeur de survivabilité du génome associé au point σ_i de l'hypercube).

Cependant, la structure de l'hypercube provoque des complications calculatoires. Afin d'éviter cette complexité, on peut supposer que les chemins évolutifs sont indépendants, et donc étudier le cas d'un arbre d'arité décroissante de N à 1 et de profondeur N , où la racine correspond au sommet σ_0 et où les $N!$ feuilles correspondent au sommet σ_1 et ont donc une survivabilité fixée à 1. Dès lors, on s'intéresse au nombre de chemins croissants dans un tel arbre, où chaque noeud possède une valeur de survivabilité aléatoire dans $[0; 1]$. Bien que différent, ce modèle fournit des résultats très proches de celui de l'hypercube (de plus, il sert d'intermédiaire de calcul, et le développement des idées dans les deux cas est similaire dans l'article de Berestycki, Brunet et Shi [BBS16]). Il constitue la base du modèle que nous allons étudier plus en détail.

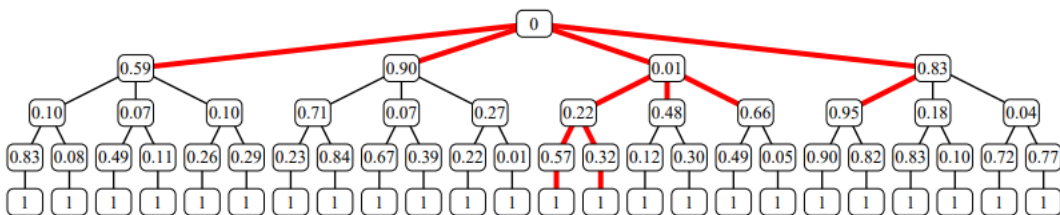


FIGURE 1.1 – Exemple d'arbre pour $N = 4$ tiré de ([BBS16]).

1.2 Le modèle de l'arbre N -aire

C'est donc l'idée de chemins croissants dans un arbre aléatoire qui reste conservée dans notre modélisation. On s'intéresse alors à un arbre N -aire, où, comme précédemment, chaque sommet σ se voit associé de manière indépendante une variable aléatoire uniforme à valeurs dans $[0; 1]$, notée x_σ . On notera alors $|\sigma|$ la profondeur du sommet σ et $\emptyset, \sigma_1, \sigma_2, \dots, \sigma_k$ la lignée ascendante du sommet σ (où $k = |\sigma|$).

L'objectif de notre étude est alors double : estimer le nombre de chemins croissants qui peuvent subsister jusqu'à une génération donnée et trouver la génération critique au-delà de laquelle aucun chemin croissant ne peut plus exister. Il s'agit évidemment d'une étude probabiliste, et l'on se posera donc ces questions dans le cadre asymptotique où $N \rightarrow \infty$. Il est important de noter que le choix d'une loi uniforme pour les variables x_σ est en réalité peu important, et que les résultats seraient similaires pour n'importe quelle loi sans atome puisque l'on s'intéresse uniquement à la croissance de la survivabilité (et non à sa valeur à chaque génération). On pourrait donc se ramener au cas uniforme par "déformation" des valeurs de survivabilité.

Pour étudier ce système, on introduit plusieurs objets et notions. Tout d'abord, un noeud σ sera dit accessible si le chemin qui le relie à la racine (notée \emptyset) est croissant, c'est à dire $x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k}$. On introduit alors, pour $k \geq 1$, $\mathcal{A}_{N,k} = \{\sigma, |\sigma| = k \text{ et } \sigma \text{ est accessible}\}$ et

$$Z_{N,k} = \#\mathcal{A}_{N,k} = \sum_{|\sigma|=k} \mathbf{1}_{\sigma \in \mathcal{A}_{N,k}} = \sum_{|\sigma|=k} \mathbf{1}_{(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k})} \quad (1.1)$$

On prêtera attention à la valeur de x_\emptyset , et on notera donc \mathbf{P}_x la mesure de probabilité $\mathbf{P}(\dots | x_\emptyset = x)$ et \mathbf{E}_x l'espérance associée. Le cas $x = 0$ étant représentatif, il sera étudié en détail. On pourra ensuite en déduire des résultats sur le cas x quelconque.

Relativement à ces objets, on s'intéressera donc aux deux aspects suivants : le comportement de $Z_{N,k}$ en fonction de k et relativement à \mathbf{P}_0 lorsque $N \rightarrow \infty$ (on sera donc amené à choisir k comme fonction de N) et les valeurs de k vérifiant $\mathbf{P}_0(Z_{N,k} \geq 1) \rightarrow 0$ lorsque $N \rightarrow \infty$ (ce qui correspond à l'extinction de la population).

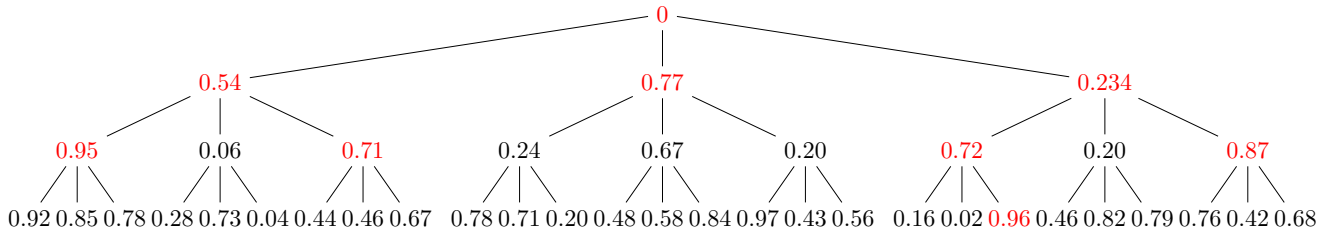


FIGURE 1.2 – Arbre 3-aire aléatoire entre les profondeurs 0 et 3.

Dans la figure 1.2 ci-dessus, les noeuds accessibles sont indiqués en rouge. La valeur de la racine est fixée à 0, et on a :

$$Z_{3,0} = 1 ; Z_{3,1} = 3 ; Z_{3,2} = 4 ; Z_{3,3} = 1$$

Dans la prochaine section, nous effectuerons quelques calculs préliminaires et révélateurs du comportement de $Z_{N,k}$, puis nous verrons par la suite deux résultats plus précis, issus principalement de l'article ([Che14]) sur le comportement de $Z_{N,k}$ lorsque $N \rightarrow \infty$: une description du comportement de la variable $Z_{N,\alpha N}$ en fonction de α ; et une approximation de la valeur d'extinction de l'arbre par $eN - 3/2 \ln(N)$. Après avoir prouvé ces résultats, nous proposerons une discrétisation du modèle afin de pouvoir réaliser quelques simulations informatiques.

Chapitre 2

Premiers calculs concernant le nombre de sommets accessibles

Dans ce chapitre, on propose une étude du moment d'ordre 1 de la variable $Z_{N,k}$, ainsi que de sa fonction génératrice. Ces calculs visent à aiguiller la réflexion qui sera menée par la suite.

2.1 Espérance du nombre de sommets accessibles

Considérons en premier lieu le moment d'ordre 1 de la variable aléatoire $Z_{N,k}$.

Tout d'abord, par indépendance des variables aléatoires $x_\emptyset, x_{\sigma_1}, \dots, x_{\sigma_k}$, on a que, pour σ fixé tel que $|\sigma| = k$,

$$\begin{aligned} \mathbf{P}(\sigma \in \mathcal{A}_{N,k}) &= \mathbf{P}(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k}) \\ &= \frac{1}{\#\mathcal{S}_{k+1}} \\ &= \frac{1}{(k+1)!} \end{aligned}$$

(on est en effet ramené au choix de la bonne permutation de sorte que $x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k}$). Il vient de même que

$$\begin{aligned} \mathbf{P}_x(\sigma \in \mathcal{A}_{N,k}) &= \mathbf{P}(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k} | x_\emptyset = x) \\ &= \mathbf{P}(x_{\sigma_1} < \dots < x_{\sigma_k}, \forall i, x_{\sigma_i} > x) \\ &= \frac{1}{k!} (1-x)^k \end{aligned}$$

Par définition de $Z_{N,k} = \sum_{|\sigma|=k} \mathbf{1}_{(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k})}$, on obtient alors :

$$\mathbf{E}_x(Z_{N,k}) = \sum_{|\sigma|=k} \mathbf{E}_x(\mathbf{1}_{(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k})}) = N^k \mathbf{P}_x(x < x_{\sigma_1} < \dots < x_{\sigma_k}) = \frac{N^k (1-x)^k}{k!} \quad (2.1)$$

En appliquant les approximations de Stirling pour $k!$, on a :

$$\forall k \geq 1, \quad 2 < \frac{k! e^k}{\sqrt{k} k^k} < 3 \quad (2.2)$$

De sorte qu'on obtient le résultat suivant :

Lemme 2.1.

$$\frac{1}{3\sqrt{k}} \left(\frac{eN(1-x)}{k} \right)^k < \mathbf{E}_x(Z_{N,k}) < \frac{1}{2\sqrt{k}} \left(\frac{eN(1-x)}{k} \right)^k \quad (2.3)$$

Ainsi, il apparaît naturel dans un premier temps de prendre k de la forme αN et d'étudier le comportement de $Z_{N,\alpha N}$ en fonction de α . On a dans ce cas (grâce à l'équivalent de Stirling) :

$$\mathbf{E}_x(Z_{N,\alpha N}) \sim_{N \rightarrow \infty} \frac{1}{\sqrt{2\pi\alpha N}} \left(\frac{e(1-x)}{\alpha} \right)^{\alpha N} \quad (2.4)$$

2.2 Fonction génératrice du nombre de sommets accessibles

Une manière classique d'envisager les questions concernant notre problème est de considérer la fonction génératrice $f_k^{(N)}$ de la variable aléatoire $Z_{N,k}$. Bien que cette méthode n'aboutisse pas, elle permet de mieux comprendre les mécanismes en jeu dans l'étude de ce système. Ainsi, posons $f_k^{(N)}(s, b) = \mathbf{E}_0(s^{Z_{N,k}(b)})$ où $Z_{N,k}(b) = \sum_{|\sigma|=k} \mathbf{1}_{(0 < x_{\sigma_1} < \dots < x_{\sigma_k} < b)}$

On peut écrire : $Z_{N,k+1}(b) = \sum_{|\omega|=1} \mathbf{1}_{x_\omega < b} \sum_{|\sigma|=k+1} \mathbf{1}_{\sigma_1=\omega} \mathbf{1}_{(x_{\sigma_1} < x_{\sigma_2} < \dots < x_{\sigma_{k+1}} < b)}$. Dès lors, il vient par indépendance des variables aléatoires attachées aux noeuds :

$$\begin{aligned} f_{k+1}^{(N)}(s, b) &= [\mathbf{E}_0(s^{\mathbf{1}_{x_\omega < b} \sum_{|\sigma|=k+1} \mathbf{1}_{\sigma_1=\omega} \mathbf{1}_{(x_{\sigma_1} < x_{\sigma_2} < \dots < x_{\sigma_{k+1}} < b)})}]^N \\ &= [1 - b + \int_0^b f_k^{(N)}(s, b - y) dy]^N \\ &= [1 - b + \int_0^b f_k^{(N)}(s, y) dy]^N \end{aligned}$$

On en déduit la relation suivante, comme $Z_{N,k}(1) = Z_{N,k}$:

$$f_{k+1}^{(N)}(s) = f_{k+1}^{(N)}(s, 1) = \left[\int_0^1 f_k^{(N)}(s, y) dy \right]^N \quad (2.5)$$

Et a fortiori :

$$\forall k \geq 2, \mathbf{P}_0(Z_{N,k} \geq 1) = 1 - f_k^{(N)}(0) = 1 - \left[\int_0^1 f_{k-1}^{(N)}(0, y) dy \right]^N \quad (2.6)$$

Cependant, de telles formules récursives sont peu exploitables en pratique pour l'étude de $Z_{N,k}$. Mais on retiendra tout de même les deux idées générales suivantes :

- L'indépendance des variables aléatoires x_σ attachées aux noeuds de l'arbre N -aire permettent de disjointre les chemins qui ont des départs différents.
- L'uniformité de la distribution des variables aléatoires x_σ permet d'effectuer des translations des valeurs sur un chemin sans changer la loi.

Chapitre 3

Résultats principaux sur l'asymptotique du nombre de chemins croissants

On établit ici les résultats principaux concernant le comportement asymptotique de $Z_{N,k}$ sous \mathbf{P}_0 , tirés de l'article [Che14]. On propose au terme de ce chapitre une adaptation au cas où x_\emptyset est initialisé à un x quelconque.

3.1 Comportement asymptotique du nombre de chemins accessibles au premier ordre

Comme le révèle l'encadrement de l'espérance de $Z_{N,k}$ (2.3), la valeur de $k = eN$ est critique dans le comportement du système. Le résultat suivant propose une description asymptotique précise de la variable $Z_{N,\alpha N}$ en fonction de la valeur de α .

Théorème 3.1. *On distingue 3 cas, et on désigne par $\theta(\alpha)$ la quantité $\alpha(1 - \ln(\alpha))$:*

1. Pour $\alpha \in]0; e[$, on a \mathbf{P}_0 - p.s. :

$$\lim_{N \rightarrow \infty} \frac{\ln(Z_{N,\alpha N})}{N} = \theta(\alpha) \quad (3.1)$$

2. Quand $N \rightarrow \infty$:

$$\mathbf{P}_0(Z_{N,eN} \geq 1) = N^{-3/2+o(1)} \quad (3.2)$$

3. Pour $\alpha > e$, on a :

$$\lim_{N \rightarrow \infty} \frac{\ln(\mathbf{P}_0(Z_{N,\alpha N} \geq 1))}{N} = \theta(\alpha) \quad (3.3)$$

Ainsi, on connaît de manière précise le comportement de $Z_{N,\alpha N}$ pour $N \rightarrow \infty$ lorsque $\alpha < e$. On donne ci-dessous le graphe de $\theta(\alpha)$:

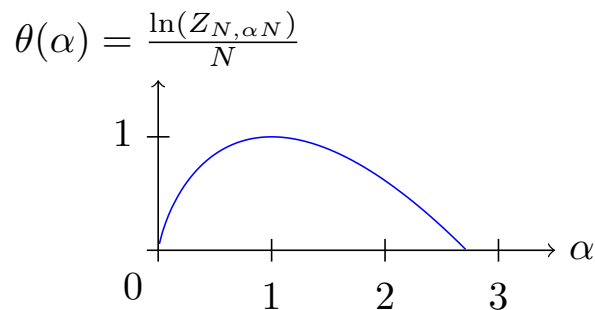


FIGURE 3.1 – Graphe de $\theta(\alpha)$ pour $0 < \alpha < e$

De plus, on observe clairement qu'au premier ordre, la génération critique d'extinction est eN , et on a même une estimation de la vitesse à laquelle $P(Z_{N,\alpha N} \geq 1)$ tend vers 0 pour $\alpha \geq e$.

3.2 Génération critique d'extinction

Le résultat suivant vient préciser ce qui précède concernant l'extinction de la variable $Z_{N,k}$. Si au premier ordre, on observe une extinction pour $k = eN$, l'étude de l'espérance de $Z_{N,eN-\beta \ln(N)}$ semble indiquer une valeur particulière en $\beta = 1/2$. En effet,

$$\begin{aligned} \mathbf{E}_0(Z_{N,eN-\beta \ln(N)}) &= \frac{N^{eN-\beta \ln(N)}}{(eN - \beta \ln(N))!} \\ &\sim \left(\frac{eN}{eN - \beta \ln(N)}\right)^{eN-\beta \ln(N)} \frac{1}{\sqrt{2\pi(eN - \beta \ln(N))}} \\ &\sim \exp(\beta \ln(N)) \frac{1}{\sqrt{2\pi N}} \end{aligned}$$

Donc finalement,

$$\mathbf{E}_0(Z_{N,eN-\beta \ln(N)}) \sim_{N \rightarrow \infty} \frac{N^{\beta-1/2}}{\sqrt{2\pi e}} \quad (3.4)$$

Cependant, une étude plus fine en probabilité aboutit à une valeur différente, présentée dans le théorème suivant :

Théorème 3.2.

$$\lim_{N \rightarrow \infty} \mathbf{P}_0(Z_{N,eN-\beta \ln(N)} \geq 1) = 1 \text{ si } \beta > 3/2 \text{ et } 0 \text{ si } \beta < 3/2 \quad (3.5)$$

Et à la valeur critique de $\beta = 3/2$, on a :

$$\forall \varepsilon > 0, \text{ pour } N \text{ assez grand, } \mathbf{P}_0(Z_{N,eN-\frac{3}{2} \ln(N)} \geq 1) \geq N^{-\varepsilon} \quad (3.6)$$

En résumé, on peut distinguer les comportements suivants pour $N \rightarrow \infty$:

	$\beta < 1/2$	$1/2 < \beta < 3/2$	$3/2 < \beta$
$\lim_{N \rightarrow \infty} \mathbf{E}_0(Z_{N,eN-\beta \ln(N)})$	0	∞	∞
$\lim_{N \rightarrow \infty} \mathbf{P}_0(Z_{N,eN-\beta \ln(N)} \geq 1)$	0	0	1

Le cas intermédiaire pour $1/2 < \beta < 3/2$ peut sembler contre-intuitif. En effet, le résultat précédent affirme qu'asymptotiquement, on a en moyenne une infinité de chemins croissants qui arrivent jusqu'à la génération $eN - \beta \ln(N)$ alors même que la probabilité qu'un tel chemin existe tend vers 0.

Il faut en fait garder en mémoire que l'espérance constitue une moyenne pondérée en probabilité. Ici, certains comportements sont très rares mais, si ils adviennent, donnent naissance à un grand nombre de chemins croissants. Il s'agit des chemins $\sigma_1, \sigma_2, \dots, \sigma_j$ où $x_{\sigma_1} < \dots < x_{\sigma_j} < \varepsilon$ pour ε très petit et $j < k$ relativement grand.

L'idée principale des démonstrations proposées en 4 est alors de ne pas considérer de tels événements lorsque l'on considère $Z_{N,k}$ et donc plutôt de se concentrer sur des chemins "réguliers" (en un sens qui sera précisé par la suite).

3.3 Adaptation au cas où la racine est initialisée à une valeur $x \neq 0$

L'intégralité des résultats précédents est présentée dans le cadre de l'étude de $Z_{N,k}$ relativement à la probabilité $\mathbf{P}_0 = \mathbf{P}(\dots | x_\emptyset = 0)$. On peut en fait transférer ces résultats à une étude selon \mathbf{P}_x . En effet, on peut tout d'abord remarquer que l'étude des chemins croissants l'arbre N-aire avec la condition $x_\emptyset = x$ se ramène à l'étude d'un arbre où chaque noeud σ possède un nombre d'enfants aléatoire suivant la loi $\mathcal{B}(N, 1-x)$ (puisque chacun des N enfants dans l'arbre originel possède une probabilité $1-x$ d'être intéressant pour notre étude).

Ainsi, on est ramené à l'étude (sans condition sur x_\emptyset) de cet objet qu'on appelle un arbre de Galton-Watson de loi de reproduction $\mathcal{B}(N, 1-x)$. La difficulté majeure pour un tel modèle est alors d'avoir accès au nombre d'individus I_k à une génération k donnée, ce qui est nécessaire pour pouvoir adapter les démonstrations des résultats des sections précédentes.

Mais nous verrons par la suite (dans la section 4) que les résultats que nous montrons sont, pour la plupart, issus d'inégalités concernant l'espérance de variables aléatoires du type $Z_{N,k} = \sum_{|\sigma|=k} \mathbf{1}_{(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k})}$. Ainsi, on rappelle le résultat classique suivant :

Lemme 3.3. *Pour des variables aléatoires indépendantes et identiquement distribuées $(X_n)_{n \geq 1}$ et une variable aléatoire I indépendante des précédentes,*

$$\mathbf{E}\left(\sum_{k=1}^I X_k\right) = \mathbf{E}(I) \times \mathbf{E}(X_1) \quad (3.7)$$

Ainsi, en notant $Z_{\mathcal{B}(N,1-x),k} = \sum_{|\sigma|=k} \mathbf{1}_{(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k})}$ le nombre de chemins croissants jusqu'à une génération k données dans l'arbre de Galton-Watson, on a :

$$\mathbf{E}_0(Z_{\mathcal{B}(N,1-x),k}) = (N(1-x))^k \mathbf{P}_0(x_\emptyset < x_{\sigma_1} < \dots < x_{\sigma_k}) \quad (3.8)$$

comme $\mathbf{E}(I_k) = N^k(1-x)^k$. Et on aura des égalités similaires pour les intermédiaires de calculs utilisés en 4.

Notons $M_N^{(x)}$ la valeur critique pour un arbre N -aire de $k = eN(1-x) - \beta \ln(N)$ au-delà de laquelle $\mathbf{P}_x(Z_{N,k} \geq 1)$ tend vers 0 pour $N \rightarrow \infty$, en écho au résultat (3.5). On obtient alors le résultat suivant :

Théorème 3.4.

$$M_N^{(x)} = M_{N(1-x)}^{(0)} = eN(1-x) - \frac{3}{2} \ln(N) \quad (3.9)$$

Chapitre 4

Preuves

Ce chapitre vise à proposer les démonstrations des résultats énoncés ci-dessus, en commençant par quelques lemmes préliminaires, mettant notamment en jeu l'idée principale de Xinxin Chen ([Che14]) : considérer les branches de l'arbre où la croissance des valeurs de survivabilité est régulière.

4.1 Lemmes

Lemme 4.1. *Inégalité de Paley-Zygmund*

Pour une variable aléatoire positive Z de variance finie, on a, pour tout $\lambda \in]0; 1[$:

$$\mathbf{P}(Z \geq \lambda \mathbf{E}(Z)) \geq (1 - \lambda)^2 \frac{\mathbf{E}(Z)^2}{\mathbf{E}(Z^2)} \quad (4.1)$$

Une telle inégalité nous sera très utile pour minorer les probabilités concernant $Z_{N,k}$ et ainsi aboutir à des résultats de convergences comme par exemple (3.6). Un grand enjeu des preuves proposées est d'appliquer ce résultat non pas à la variable $Z_{N,k}$ en tant que telle mais à des variables aléatoires annexes. Une telle variable aléatoire $Z < Z_{N,k}$ sera choisie pour que $\mathbf{E}(Z)$ ait une forme plus facilement exploitable et que $\mathbf{E}(Z^2)$ soit majoré par une expression de la forme $c\mathbf{E}(Z)$ ou $c\mathbf{E}(Z)^2$ (avec c une constante positive).

Démonstration. En effet, $Z = Z\mathbf{1}_{Z < \lambda \mathbf{E}(Z)} + Z\mathbf{1}_{Z \geq \lambda \mathbf{E}(Z)}$. Naturellement, $\mathbf{E}(Z\mathbf{1}_{Z < \lambda \mathbf{E}(Z)}) \leq \lambda \mathbf{E}(Z)$ et par inégalité de Cauchy-Schwarz, $\mathbf{E}(Z\mathbf{1}_{Z \geq \lambda \mathbf{E}(Z)})^2 \leq \mathbf{E}(Z^2)\mathbf{P}(Z \geq \lambda \mathbf{E}(Z))$.

On en déduit que $\mathbf{P}(Z \geq \lambda \mathbf{E}(Z)) \geq \frac{\mathbf{E}(Z - Z\mathbf{1}_{Z < \lambda \mathbf{E}(Z)})^2}{\mathbf{E}(Z^2)} \geq (1 - \lambda)^2 \frac{\mathbf{E}(Z)^2}{\mathbf{E}(Z^2)}$ □

Comme introduit précédemment, l'idée principale des preuves proposées est la suivante : il suffit de considérer certains types de chemins (que l'on pourrait qualifier de "réguliers" car la croissance des valeurs des variables aléatoires est contrôlée inférieurement), qui permettent d'approcher de manière satisfaisante la variable aléatoire $Z_{N,k}$, tout en permettant des calculs plus faciles. L'intuition de considérer des chemins "réguliers" dans l'arbre provient de l'observation suivante :

$$\mathbf{E}(x_j | x_1 < x_2 < \dots < x_k) = \frac{j}{k+1}$$

pour des variables aléatoires x_1, \dots, x_k indépendantes et à loi uniforme dans $[0; 1]$. Ce résultat découle d'un calcul intégral direct.

Lemme 4.2. Pour x_i des variables aléatoires indépendantes identiquement distribuées de loi uniforme dans $[0; 1]$, on a :

$$\forall 1 \leq k \leq J-1, \varphi(k, J) = \mathbf{P}(x_1 < x_2 < \dots < x_k, \forall 1 \leq j \leq k, x_j \geq \frac{j}{J}) = \frac{J-k}{k!J} \quad (4.2)$$

Mais aussi, $\forall \varepsilon \in [0; 1], \forall 1 \leq k \leq J$:

$$\begin{aligned} \psi(k, J, \varepsilon) &= \mathbf{P}(x_1 < x_2 < \dots < x_k, \forall 1 \leq j \leq k, x_j \geq \varepsilon + (1 - \varepsilon)\frac{j-1}{J}) \\ &= \frac{(1 + 1/J)^k (J + 1 - k)}{k!(J + 1)} (1 - \varepsilon)^k \end{aligned} \quad (4.3)$$

On notera alors $\mathcal{A}_{N,k,\varepsilon} = \{\sigma, |\sigma| = k, x_{\sigma_1} < \dots < x_{\sigma_k}, \forall 1 \leq j \leq k, x_j \geq \varepsilon + (1 - \varepsilon)\frac{j-1}{k}\}$ et $Z_{N,k,\varepsilon} = \#\mathcal{A}_{N,k,\varepsilon}$. Il vient alors :

$$\mathbf{P}_0(\sigma \in \mathcal{A}_{N,k,\varepsilon}) = \psi(k, k, \varepsilon) \quad (4.4)$$

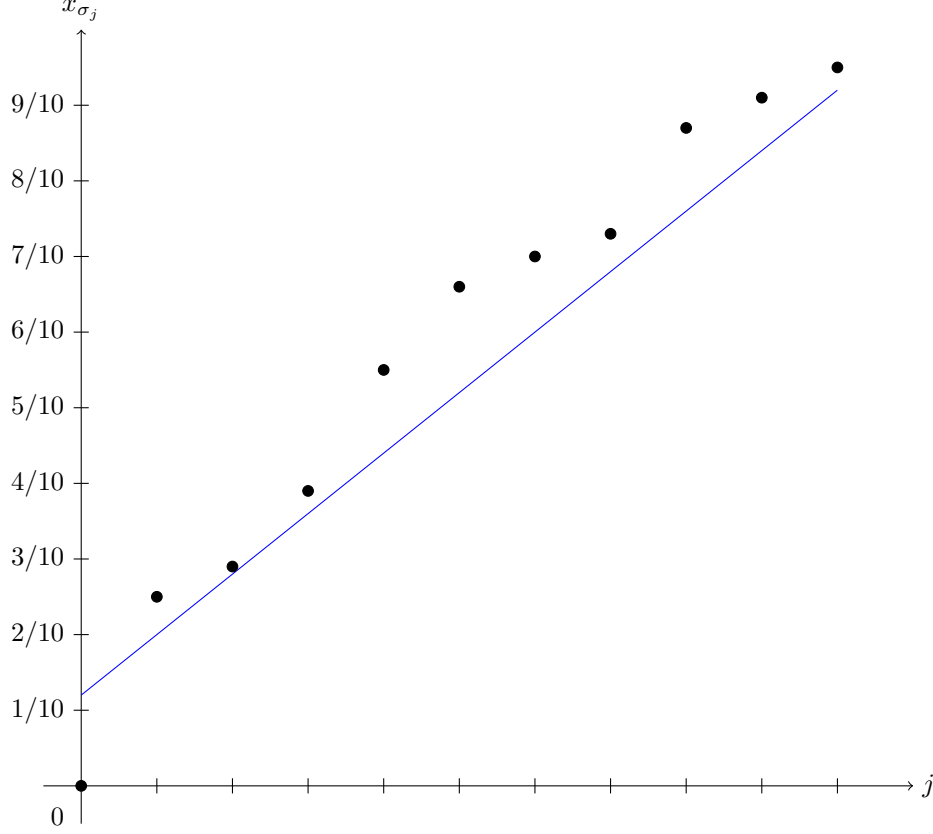


FIGURE 4.1 – Exemple de chemin dans $\mathcal{A}_{N,k,\varepsilon}$ avec $k = 10$ et $\varepsilon = 1/5$

Démonstration. La preuve du premier résultat se fait par un calcul direct :

$$\varphi(k, J) = \int_{k/J}^1 \int_{(k-1)/J}^{x_k} \dots \int_{1/J}^{x_2} dx_1 \dots dx_{k-1} dx_k = \frac{j}{J} = \frac{J-k}{k!J}$$

Pour prouver le second, on observe (en exprimant $\psi(k, J, \varepsilon)$ sous forme intégrale et par un changement de variables) que $\psi(k, J, \varepsilon) = \psi(k, J, 0)(1 - \varepsilon)^k$. En remarquant que $\psi(k, J, \frac{1}{J+1}) = \varphi(k, J+1)$, on peut donc conclure. \square

Considérons maintenant un autre type de chemins particuliers. On définit l'ensemble suivant pour $1 \leq L \leq K$:

$$A_L(K) = \{x_1 < x_2 < \dots < x_K, \forall 1 \leq j \leq K, x_j \geq \frac{(j-L)_+}{K+1}\} \quad (4.5)$$

avec les x_j des variables aléatoires uniformes dans $[0; 1]$.

On a alors la majoration fondamentale suivante portant sur l'ensemble $A_L(K)$:

Lemme 4.3.

$$\exists c_0 > 0, \forall 1 \leq L \leq K, \mathbf{P}(A_L(K)) \leq \frac{e^{c_0 \sqrt{L}} e^K}{K^{3/2} (K+1)^K} \quad (4.6)$$

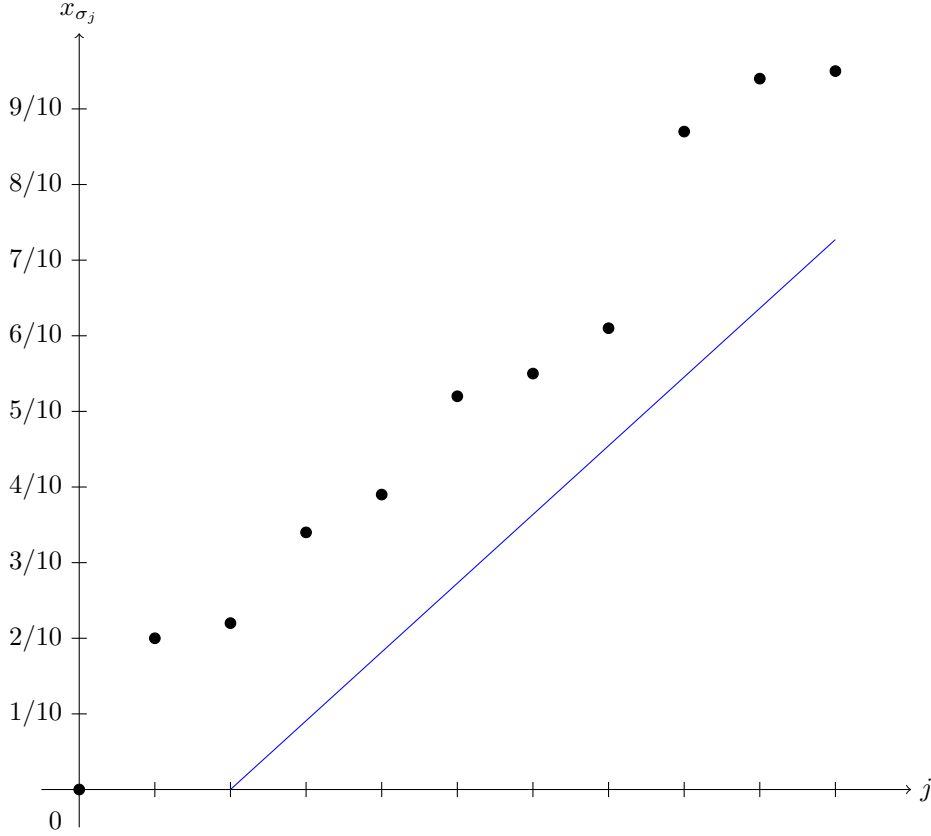


FIGURE 4.2 – Exemple de chemin dans $\mathcal{A}_L(k)$ avec $k = 10$ et $L = 2$

Idées de preuve : Cette preuve très calculatoire repose sur des raisonnements par récurrence. On en présente ici les principales idées mais les détails ne seront pas présentés (on peut se référer à [Che14]).

Tout d'abord, pour $L = 1$, on obtient (en utilisant (2.2)) :

$$\mathbf{P}(A_L(K)) = \psi(K, K, 0) = \frac{(1 + 1/K)^K}{(K+1)!} \leq \frac{e^{2\sqrt{L}}}{K^{3/2}} \frac{e^K}{(K+1)^K}$$

Ensuite, $A_1(K) \subset A_2(K) \subset \dots \subset A_L(K)$, donc on se ramène à estimer $\mathbf{P}(A_{i+1}(K) - A_i(K))$. Or, on a :

$$\begin{aligned} & A_{i+1}(K) - A_i(K) \\ &= \{x_1 < x_2 < \dots < x_K, \forall 1 \leq j \leq K, x_j \geq \frac{(j-i-1)_+}{K+1}, \exists 1 \leq j \leq K, x_j < \frac{(j-i)_+}{K+1}\} \\ &\subset \bigsqcup_{k=i+1}^K \{x_1 < x_2 < \dots < x_K, \forall i+1 \leq j \leq k-1, x_j \geq \frac{j-i}{K+1}, x_k < \frac{k-i}{K+1}, \forall k+1 \leq j \leq K, x_j < \frac{j-i-1}{K+1}\} \\ &= \bigsqcup_{k=i+1}^K C_{i,k} \end{aligned}$$

Puis, par indépendance des variables aléatoires x_i , on a $\mathbf{P}(C_{i,k}) = p_{i,k}q_{i,k}$ où :

$$\begin{aligned} p_{i,k} &= \mathbf{P}(\{x_1 < x_2 < \dots < x_k \frac{k-i}{K+1}, \forall i+1 \leq j \leq k, x_j \geq \frac{j-i}{K+1}\}) \\ q_{i,k} &= \mathbf{P}(\{x_{k+1} < \dots < x_K, \forall k+1 \leq j \leq K, x_j \geq \frac{j-i-1}{K+1}\}) \end{aligned}$$

Finalement, $q_{i,k}$ se calcule explicitement avec (4.3) tandis que $p_{i,k}$ est majoré par un habile raisonnement par récurrence.

4.2 Théorème de convergence de $Z_{N,\alpha N}$

Preuve du cas $0 < \alpha < e$ (3.1) :

On veut montrer que $\mathbf{P}_0 - p.s., \lim_{N \rightarrow \infty} \frac{\ln(Z_{N,\alpha N})}{N} = \theta(\alpha)$.

Tout d'abord, par la majoration de l'espérance obtenue en 2.1, on a que :

$$\mathbf{E}_0(Z_{N,\alpha N}) \leq \frac{1}{2\sqrt{\alpha N}} \left(\frac{e}{\alpha}\right)^{\alpha N} = \frac{1}{2\sqrt{\alpha N}} e^{\theta(\alpha)N}$$

En appliquant l'inégalité de Markov pour $\delta > 0$, il vient :

$$\mathbf{P}_0(Z_{N,\alpha N} \geq \exp((\theta(\alpha) + \delta)N)) \leq \frac{e^{-\delta N}}{2\sqrt{\alpha N}}$$

Dès lors, par le lemme de Borel-Cantelli, on a que $\forall \delta > 0$:

$$\mathbf{P}_0(\limsup_{N \rightarrow \infty} \frac{\ln(Z_{N,\alpha N})}{N} \geq \theta(\alpha) + \delta) = 0 \quad (4.7)$$

D'où une majoration \mathbf{P}_0 presque-sûrement.

On souhaite désormais obtenir une minoration $\mathbf{P}_0 - p.s.$ de $\liminf_{N \rightarrow \infty} \frac{\ln(Z_{N,\alpha N})}{N}$. Pour cela, on remarque tout d'abord que :

$$Z_{N,\alpha N} \geq \sum_{|\omega|=1} \mathbf{1}_{x_\omega < \varepsilon} \sum_{|\sigma|=\alpha N} \mathbf{1}_{\sigma \in \mathcal{B}_\omega}$$

où $\mathcal{B}_\omega = \{\sigma, |\sigma| = \alpha N, \sigma_1 = \omega, \varepsilon < x_{\sigma_2} < \dots < x_{\sigma_{\alpha N}} = x_\sigma\}$ Dès lors, pour tout C , par indépendance des variables aléatoires attachées aux sommets,

$$\mathbf{P}_0(Z_{N,\alpha N} < C) \leq \mathbf{P}_0\left(\sum_{|\omega|=1} \mathbf{1}_{x_\omega < \varepsilon} \sum_{|\sigma|=\alpha N} \mathbf{1}_{\sigma \in \mathcal{B}_\omega} < C\right) = [\mathbf{P}_0(\mathbf{1}_{x_\omega < \varepsilon} \sum_{|\sigma|=\alpha N} \mathbf{1}_{\sigma \in \mathcal{B}_\omega} < C)]^N$$

Et donc, en distinguant le cas $x_\omega \geq \varepsilon$ et $x_\omega < \varepsilon$, et en conditionnant par le deuxième événement, on obtient :

$$\mathbf{P}_0(Z_{N,\alpha N} < C) \leq [(1 - \varepsilon) + \varepsilon \mathbf{P}_0\left(\sum_{|\sigma|=\alpha N} \mathbf{1}_{\sigma \in \mathcal{B}_\omega} < C | x_\omega < \varepsilon\right)]^N$$

Comme $\mathbf{P}_0(\sigma \in \mathcal{B}_\omega | x_\omega < \varepsilon) = \mathbf{P}_\varepsilon(\sigma \in \mathcal{A}_{N,\alpha N-1})$, on a finalement que :

$$\mathbf{P}_0(Z_{N,\alpha N} < C) \leq [(1 - \varepsilon) + \varepsilon \mathbf{P}_\varepsilon(Z_{N,\alpha N-1} < C)]^N = [1 - \varepsilon \mathbf{P}_\varepsilon(Z_{N,\alpha N-1} \geq C)]^N \quad (4.8)$$

Or, par définition de $Z_{N,\alpha N-1,\varepsilon}$ (donnée en (4.4)), on a $\mathbf{P}_\varepsilon(Z_{N,\alpha N-1} \geq C) \geq \mathbf{P}_0(Z_{N,\alpha N-1,\varepsilon} \geq C)$ On va donc appliquer la formule précédente avec $C = \exp(\theta(\alpha)N - 3\alpha\varepsilon N) \leq \frac{1}{2} \mathbf{E}_0(Z_{N,\alpha N-1,\varepsilon})$ lorsque N est grand et ε suffisamment petit (dépendant de α). Dans ce cas, on a (avec (4.1)) :

$$\mathbf{P}_\varepsilon(Z_{N,\alpha N-1} \geq C) \geq \mathbf{P}_0(Z_{N,\alpha N-1,\varepsilon} \geq \frac{1}{2} \mathbf{E}_0(Z_{N,\alpha N-1,\varepsilon})) \geq \frac{1}{4} \frac{\mathbf{E}_0(Z_{N,\alpha N-1,\varepsilon})^2}{\mathbf{E}_0(Z_{N,\alpha N-1,\varepsilon}^2)}$$

Or, en notant $\sigma \wedge \sigma'$ le plus profond ancêtre commun des noeuds σ et σ' ,

$$\begin{aligned} \mathbf{E}_0(Z_{N,k,\varepsilon}^2) &= \mathbf{E}_0(Z_{N,k,\varepsilon}) + \mathbf{E}_0\left(\sum_{q=0}^{k-1} \sum_{|\sigma \wedge \sigma'|=q} \mathbf{1}_{\sigma, \sigma' \in \mathcal{A}_{N,k,\varepsilon}}\right) \\ &= \mathbf{E}_0(Z_{N,k,\varepsilon}) + \sum_{q=0}^{k-1} N^q N(N-1) N^{2k-2q-2} \mathbf{P}_0(\sigma, \sigma' \in \mathcal{A}_{N,k,\varepsilon} | |\sigma \wedge \sigma'| = q) \end{aligned}$$

Puis, on a la majoration suivante, d'après (4.4) :

$$\begin{aligned} \mathbf{P}_0(\sigma, \sigma' \in \mathcal{A}_{N,k,\varepsilon} | |\sigma \wedge \sigma'| = q) &= \int_{\varepsilon+(1-\varepsilon)(q-1)/k}^1 \mathbf{P}_0(\sigma, \sigma' \in \mathcal{A}_{N,k} | |\sigma \wedge \sigma'| = q, x_{\sigma \wedge \sigma'} = y) dy \\ &\leq \int_{\varepsilon+(1-\varepsilon)(q-1)/k}^1 \mathbf{P}_0(\sigma \in \mathcal{A}_{N,k,\varepsilon} | x_{\sigma_q} = y) \times \mathbf{P}_0(\sigma \in \mathcal{A}_{N,k-q,\varepsilon+(1-\varepsilon)q/k}) dy \\ &= \psi(k, k, \varepsilon) \psi(k-q, k-q, \varepsilon + (1-\varepsilon)q/k) \end{aligned}$$

Comme $\mathbf{E}_0(Z_N, k, \varepsilon) = N^k \psi(k, k, \varepsilon) \geq 1$ pour $k = \alpha N - 1$ et N suffisamment grand, on obtient lorsque $k = \alpha N - 1$:

$$\mathbf{E}_0(Z_{N,k,\varepsilon}^2) \leq \mathbf{E}_0(Z_{N,k,\varepsilon})^2 \left(1 + \sum_{q=0}^{k-1} \frac{N-1}{N} N^{-q} \frac{\psi(k-q, k-q, \varepsilon + (1-\varepsilon)q/k)}{\psi(k, k, \varepsilon)}\right) \leq \mathbf{E}_0(Z_{N,k,\varepsilon})^2 (1+c) \quad (4.9)$$

Avec c une constante strictement positive. (Cette majoration utilise juste l'expression explicite de ψ donnée en (4.3)).

En réinjectant dans (4.8), avec $C = \exp(\theta(\alpha)N - 3\alpha\varepsilon N)$, on obtient :

$$\mathbf{P}_0(Z_{N,\alpha N} \geq \exp(\theta(\alpha)N - 3\alpha\varepsilon N)) \leq e^{-N\varepsilon \frac{1}{4(1+c)}}$$

Et on peut donc conclure par le lemme de Borel-Cantelli que $\forall \varepsilon > 0$ suffisamment petit :

$$\mathbf{P}_0(\liminf_{N \rightarrow \infty} \frac{\ln(Z_{N,\alpha N})}{N} \leq \theta(\alpha) - 3\alpha\varepsilon) = 0 \quad (4.10)$$

Preuve du cas $\alpha > e$ (3.3) :

On montre ici que $\lim_{N \rightarrow \infty} \frac{\ln(\mathbf{P}_0(Z_{N,\alpha N} \geq 1))}{N} = \theta(\alpha)$.

Tout d'abord, par inégalité de Markov et d'après la majoration obtenue en 2.1, on obtient :

$$\mathbf{P}_0(Z_{N,\alpha N} \geq 1) \leq \mathbf{E}_0(Z_{N,\alpha N}) \leq \frac{e^{\theta(\alpha)N}}{2\sqrt{\alpha N}}$$

Donc :

$$\limsup_{N \rightarrow \infty} \frac{\ln(\mathbf{P}_0(Z_{N,\alpha N} \geq 1))}{N} \leq \theta(\alpha) \quad (4.11)$$

Pour minorer, on utilise l'idée fondamentale de l'article en considérant la variable $Z_{N,\alpha N,0} \leq Z_{N,\alpha N}$. Dès lors par inégalité de Cauchy-Scharz, il vient :

$$\mathbf{P}_0(Z_{N,\alpha N} \geq 1) \geq \mathbf{P}_0(Z_{N,\alpha N,0} \geq 1) \geq \frac{\mathbf{E}_0(Z_{N,\alpha N,0})^2}{E_0(Z_{N,\alpha N,0}^2)}$$

Or, de même que pour (4.9), on a une constante $c > 0$ telle que :

$$\mathbf{E}_0(Z_{N,k,0}) \leq \mathbf{E}_0(Z_{N,k,0}) \left(1 + \sum_{q=0}^{k-1} N^{k-q} \psi(k-q, k-q, q/k)\right) \leq \mathbf{E}_0(Z_{N,k,0}) (1+c)$$

Or, d'après (4.3) et les approximations de Stirling (2.2),

$$\psi(\alpha N, \alpha N, 0) = \frac{(\alpha N + 1)^{\alpha N}}{(\alpha N)^{\alpha N} (\alpha N + 1)!} \geq \frac{e^{\alpha N}}{3(\alpha N + 1)^{3/2} (\alpha N)^{\alpha N}}$$

Donc on obtient :

$$\mathbf{P}_0(Z_{N,\alpha N} \geq 1) \geq \frac{1}{1+c} \mathbf{E}_0(Z_{N,\alpha N,0}) = \frac{1}{1+c} N^{\alpha N} \psi(\alpha N, \alpha N, 0) \geq \frac{e^{\theta(\alpha)N}}{3(c+1)(\alpha N + 1)^{3/2}} \quad (4.12)$$

D'où :

$$\liminf_{N \rightarrow \infty} \frac{\ln(\mathbf{P}_0(Z_{N,\alpha N} \geq 1))}{N} \geq \theta(\alpha) \quad (4.13)$$

On peut donc conclure au résultat.

Preuve du cas $\alpha = e$ (3.2) :

On montre ici que pour $N \rightarrow \infty$, $\mathbf{P}_0(Z_{N,eN} \geq 1) = N^{-3/2+o(1)}$.

Tout d'abord, on peut observer que les calculs menés pour aboutir à (4.12) restent valables pour $\alpha = e$ de sorte qu'on obtient une constante λ telle que :

$$\mathbf{P}_0(Z_{N,eN} \geq 1) \geq \lambda N^{-3/2} \quad (4.14)$$

Pour majorer, on utilise l'événement $\mathcal{A}_L(k)$ défini en (4.5) pour $k = eN$ et $L = 2 \ln(k)$. Puisque :

$$\mathcal{A}_{N,k} \subset \mathcal{A}_L(k) \cup \bigcup_{i=L+1}^k \{ \exists |\sigma| = i, x_{\sigma_1} < \dots < x_{\sigma_i}, x_{\sigma_i} < \frac{i-L}{k+1} \}$$

On obtient :

$$\mathbf{P}_0(Z_{N,eN} \geq 1) \leq \mathbf{P}_0(\exists \sigma \in \mathcal{A}_L(k)) + \sum_{i=L+1}^k \mathbf{P}_o(\exists |\sigma| = i, x_{\sigma_1} < \dots < x_{\sigma_i}, x_{\sigma_i} < \frac{i-L}{k+1}) \quad (4.15)$$

D'une part par inégalité de Markov et d'après le lemme (4.6),

$$\mathbf{P}_0(\exists \sigma \in \mathcal{A}_L(k)) \leq \mathbf{E}_0\left(\sum_{|\sigma|=k} \mathbf{1}_{(\sigma \in \mathcal{A}_L(k))}\right) \leq N^k \frac{e^{c_0 \sqrt{L}} e^k}{k^{3/2} (k+1)^k} = O\left(\frac{e^{c_0 \sqrt{2 \ln(eN)}}}{N^{3/2}}\right)$$

D'autre part, en appliquant aussi l'inégalité de Markov :

$$\begin{aligned} \sum_{i=L+1}^k \mathbf{P}_o(\exists |\sigma| = i, x_{\sigma_1} < \dots < x_{\sigma_i}, x_{\sigma_i} < \frac{i-L}{k+1}) &\leq \sum_{i=L+1}^k \mathbf{E}_0\left(\sum_{\sigma \in \mathcal{A}_{N,i}} \mathbf{1}_{(x_{\sigma_i} < \frac{i-L}{k+1})}\right) \\ &= \sum_{i=L+1}^k N^i \left(\frac{i-L}{k+1}\right)^i \frac{1}{i!} \\ &\leq \sum_{i=L+1}^k \left(\frac{eN}{k+1}\right)^i \frac{e^{-L}}{2\sqrt{i}} \\ &= O(N^{-3/2}) \end{aligned}$$

(En utilisant le fait que $\sum_1^k \frac{1}{\sqrt{i}} = O(\sqrt{k})$)

En combinant les deux majorations dans (4.15), on obtient donc :

$$\mathbf{P}_0(Z_{N,eN} \geq 1) \leq N^{-3/2+o(1)} \quad (4.16)$$

4.3 Estimation asymptotique du temps d'extinction $M_N^{(0)}$ au second ordre

Preuve du cas $\beta < 3/2$ (3.5) :

Fixons $\beta < 3/2$ et $k = eN - \beta \ln N$. Alors, de même qu'en (4.15), on obtient :

$$\mathbf{P}_0(Z_{N,k} \geq 1) \leq N^k \mathbf{P}(\mathcal{A}_L(k)) + \sum_{i=L+1}^k \left(\frac{eN}{k+1}\right)^i \frac{e^{-L}}{2\sqrt{i}} \quad (4.17)$$

Or, avec $L = 2 \ln(N)$, il vient :

$$\begin{aligned} \sum_{i=L+1}^k \left(\frac{eN}{k+1}\right)^i \frac{e^{-L}}{2\sqrt{i}} &\leq \sum_{i=L+1}^k \left(1 + \frac{\beta \ln(N)}{K}\right)^i \frac{N^{-2}}{2\sqrt{i}} \\ &\leq \sum_{i=L+1}^k \frac{N^{\beta-2}}{2\sqrt{i}} \\ &= O(N^{\beta-3/2}) \end{aligned}$$

Et aussi via (4.6) :

$$N^k \mathbf{P}(\mathcal{A}_L(k)) \leq \frac{e^{c_0 \sqrt{L}}}{k^{3/2}} \left(\frac{eN}{k+1}\right)^k \leq \frac{e^{c_0 \sqrt{L}}}{k^{3/2}} \left(1 + \frac{\beta \ln(N)}{k}\right)^k = O(N^{\beta-3/2} e^{c_0 \sqrt{L}})$$

En combinant ces deux estimées dans (4.17), on obtient bien que pour $\beta < 3/2$:

$$\lim_{N \rightarrow \infty} \mathbf{P}(Z_{N,eN-\beta \ln(N)} \geq 1) = 0 \quad (4.18)$$

Preuve du cas $\beta > 3/2$ (3.5) :

On veut ici prouver que $\lim_{N \rightarrow \infty} \mathbf{P}(Z_{N, eN - \beta \ln(N)} \geq 1) = 1$ pour $\beta > 3/2$.

Pour cela, on pose $k_0 = \gamma \ln(N)$, $K_0 = eN - (\beta + \gamma) \ln(N)$ et $\delta_N = \frac{(\gamma + \beta) \ln(N)}{eN}$. On s'intéresse alors aux noeuds σ tels que :

$$0 < x_{\sigma_1} < \dots < x_{\sigma_{k_0}} < \delta_N; \delta_N < x_{\sigma_{k_0+1}} < \dots < x_{\sigma_{k_0+K_0}} \text{ et } x_{\sigma_{k_0+j}} \geq \delta_N + (1 - \delta_N) \frac{j-1}{K_0}$$

Naturellement, en notant \mathcal{E} l'ensemble de tels noeuds accessibles, on a que $\mathbf{P}_0(Z_{N, eN - \beta \ln(N)} = 0) \leq \mathbf{P}_0(\#\mathcal{E} = 0)$, et on obtient donc, en coupant au rang k_0 :

$$\begin{aligned} \mathbf{P}_0(Z_{N, eN - \beta \ln(N)} = 0) &\leq \mathbf{E}_0(\mathbf{1}_{\#\mathcal{E}=0}) \\ &= \mathbf{E}_0((\mathbf{1}_{Z_{N, K_0, \delta_N} = 0})^{Z_{N, k_0}(\delta_N)}) \\ &= \mathbf{P}_0((\mathbf{1}_{Z_{N, K_0, \delta_N} = 0})^{Z_{N, k_0}(\delta_N)} = 1) \\ &= \sum_{z=0}^{k_0} \mathbf{P}(Z_{N, k_0}(\delta_N) = z) \mathbf{P}_0(\mathbf{1}_{Z_{N, K_0, \delta_N} = 0} = 1)^z \\ &= \mathbf{E}_0(\mathbf{P}_0(Z_{N, K_0, \delta_N} = 0)^{Z_{N, k_0}(\delta_N)}) \\ &= \mathbf{E}_0((1 - \mathbf{P}_0(Z_{N, K_0, \delta_N} \geq 1))^{Z_{N, k_0}(\delta_N)}) \end{aligned} \quad (4.19)$$

Tout d'abord, on cherche donc à minorer $\mathbf{P}_0(Z_{N, K_0, \delta_N} \geq 1)$. Par inégalité de Cauchy-Schwarz, on a :

$$\mathbf{P}_0(Z_{N, K, \delta} \geq 1) \geq \frac{\mathbf{E}_0(Z_{N, K, \delta})^2}{\mathbf{E}_0(Z_{N, K, \delta}^2)} \quad (4.20)$$

Or, par (4.4), on a pour $\delta = \delta_N$ et $K = K_0$:

$$aN^{-3/2} \leq \mathbf{E}_0(Z_{N, K, \delta}) = N^K \frac{(1 + 1/K)^K}{(K + 1)!} (1 - \delta)^K \leq bN^{-3/2}$$

où a et b sont des constantes positives.

De plus, de même qu'en (4.9), on a $C, C' > 0$ telles que :

$$\begin{aligned} \mathbf{E}_0(Z_{N, K, \delta}^2) &\leq \mathbf{E}_0(Z_{N, K, \delta})(1 + \mathbf{E}_0(Z_{N, K, \delta}) \sum_{q=0}^{K-1} \frac{N-1}{N} N^{-q} \frac{\psi(K-q, K-q, \delta + (1-\delta)q/K)}{\psi(K, K, \delta)}) \\ &\leq \mathbf{E}_0(Z_{N, K, \delta})(1 + \mathbf{E}_0(Z_{N, K, \delta}) C \sum_{q=0}^{K-1} (\frac{K}{K-q})^{3/2} (\frac{K}{e(1-\delta)K})^q) \\ &\leq \mathbf{E}_0(Z_{N, K, \delta})(1 + \mathbf{E}_0(Z_{N, K, \delta}) C' N^{3/2}) \\ &\leq \mathbf{E}_0(Z_{N, K, \delta})(1 + bC') \end{aligned}$$

(en prenant $K = K_0$ et $\delta = \delta_N$) Finalement, en réinjectant dans (4.20), on obtient $\lambda > 0$ de sorte que :

$$\mathbf{P}_0(Z_{N, K_0, \delta_N} \geq 1) \geq \lambda N^{-3/2} \quad (4.21)$$

Et donc,

$$\mathbf{P}_0(Z_{N, eN - \beta \ln(N)} = 0) \leq \mathbf{E}_0((1 - \lambda N^{-3/2})^{Z_{N, k_0}(\delta_N)}) \quad (4.22)$$

Il s'agit désormais de prouver que $Z_{N, k_0}(\delta_N)$ est grand avec une probabilité élevée (dans un sens que nous allons préciser). On utilise pour ce faire deux paramètres ε_N et δ'_N de sorte que $\varepsilon_N + \delta'_N = \delta_N$. Dans ce cas, il vient pour ε tel que $\beta - 3\varepsilon > 3/2$:

$$\begin{aligned} Z_{N, k_0}(\delta_N) &= \sum_{|\sigma|=k_0} \mathbf{1}_{(0 < x_{\sigma_1} < \dots < x_{\sigma_{k_0}} < \delta_N)} \\ &\leq \sum_{|\omega|=1} \mathbf{1}_{x_\omega < \varepsilon_N} \sum_{|\sigma|=k_0; \sigma_1 = \omega} \mathbf{1}_{(\varepsilon_N < x_{\sigma_2} < \dots < x_{\sigma_{k_0}} < \delta_N = \varepsilon_N + \delta'_N; \forall j \geq 2, x_{\sigma_j} \geq \varepsilon_N + \frac{j-2}{k_0-1} \delta'_N)} \end{aligned}$$

Or, quitte à traduire tous les noeuds par $1 - \delta_N = 1 - \delta'_N - \varepsilon_N$, on observe que les variables aléatoires $\sum_{|\sigma|=k_0; \sigma_1=\omega} \mathbf{1}_{(\varepsilon_N < x_{\sigma_2} < \dots < x_{\sigma} < \delta_N = \varepsilon_N + \delta'_N; \forall j \geq 2, x_{\sigma_j} \geq \varepsilon_N + \frac{j-2}{k_0-1} \delta'_N)}$ et $Z_{N, k_0-1, 1-\delta'_N}$ ont même distribution de probabilité sous \mathbf{P}_0 de sorte que :

$$\begin{aligned} \mathbf{P}_0(Z_{N, k_0}(\delta_N) \leq N^{\frac{3+\varepsilon}{2}}) &\leq (\mathbf{P}_0(\mathbf{1}_{x_\omega < \varepsilon_N} \sum_{|\sigma|=k_0; \sigma_1=\omega} \mathbf{1}_{(\varepsilon_N < x_{\sigma_2} < \dots < x_{\sigma} < \delta_N = \varepsilon_N + \delta'_N; \forall j \geq 2, x_{\sigma_j} \geq \varepsilon_N + \frac{j-2}{k_0-1} \delta'_N)} \leq N^{\frac{3+\varepsilon}{2}}))^{N^l} \\ &= (1 - \varepsilon_N + \varepsilon_N \mathbf{P}_0(Z_{N, k_0-1, 1-\delta'_N} \leq N^{\frac{3+\varepsilon}{2}}))^{N^l} \end{aligned} \quad (4.23)$$

Or, avec $\delta'_N = \frac{(\beta - \varepsilon + \gamma) \ln(N)}{\varepsilon_N}$, on obtient que :

$$\mathbf{E}_0(Z_{N, k_0-1, 1-\delta'_N}) \geq 2N^{\frac{3+\varepsilon}{2}}$$

lorsque N et γ sont suffisamment grands. On a aussi, de même qu'en (4.9) qu'il existe une constante $C(\beta, \gamma) > 0$ telle que :

$$\mathbf{E}_0(Z_{N, k_0-1, 1-\delta'_N}^2) \leq C(\beta, \gamma) \mathbf{E}_0(Z_{N, k_0-1, 1-\delta'_N})^2$$

Finalement, par inégalité de Paley-Zygmund (4.1), on a :

$$\mathbf{P}_0(Z_{N, k_0-1, 1-\delta'_N} \geq N^{\frac{3+\varepsilon}{2}}) \geq \mathbf{P}_0(Z_{N, k_0-1, 1-\delta'_N} \geq \frac{1}{2} \mathbf{E}_0(Z_{N, k_0-1, 1-\delta'_N})) \geq \frac{1}{4C(\beta, \gamma)} = l > 0 \quad (4.24)$$

En réinjectant dans (4.23), il vient donc :

$$\begin{aligned} \mathbf{P}_0(Z_{N, k_0}(\delta_N) \leq N^{\frac{3+\varepsilon}{2}}) &\leq (1 - \varepsilon_N + \varepsilon_N (1 - \mathbf{P}_0(Z_{N, k_0-1, 1-\delta'_N} \geq N^{\frac{3+\varepsilon}{2}})))^{N^l} \\ &= \exp(N \ln(1 - \varepsilon_N \mathbf{P}_0(Z_{N, k_0-1, 1-\delta'_N} \geq N^{\frac{3+\varepsilon}{2}}))) \\ &\leq \exp(-N \varepsilon_N \mathbf{P}_0(Z_{N, k_0-1, 1-\delta'_N} \geq N^{\frac{3+\varepsilon}{2}})) \\ &\leq e^{-N \varepsilon_N l} \end{aligned}$$

qui tend vers 0 quand $N \rightarrow \infty$ car $\varepsilon_N = \frac{\varepsilon \ln(N)}{e^N}$.

En reprenant depuis (4.22), on obtient finalement la convergence souhaitée :

$$\begin{aligned} \mathbf{P}_0(Z_{N, eN-3/2 \ln(N)} = 0) &\leq \mathbf{E}_0((1 - \lambda N^{-3/2})^{Z_{N, k_0}(\delta_N)}; Z_{N, k_0}(\delta_N) \geq N^{\frac{3+\varepsilon}{2}}) + \mathbf{P}_0(Z_{N, k_0}(\delta_N) \leq N^{\frac{3+\varepsilon}{2}}) \\ &\leq (1 - \lambda N^{-3/2})^{N^{\frac{3+\varepsilon}{2}}} + e^{-N \varepsilon_N l} \rightarrow 0 \end{aligned} \quad (4.25)$$

lorsque $N \rightarrow \infty$.

Preuve du cas $\beta = 3/2$ (3.6) :

De même que dans la section précédente, on a, pour $\beta = 3/2$ et pour tout $\varepsilon > 0$, que :

$$\begin{aligned} \mathbf{P}_0(Z_{N, eN-3/2 \ln(N)} \geq 1) &\geq \mathbf{P}_0(\#\mathcal{E} \geq 1) \\ &= \mathbf{P}_0\left(\sum_{|\omega|=k_0} \mathbf{1}_{(x_{\omega_1} < \dots < x_{\omega} < \delta_N)} \sum_{|\sigma|=k_0+K_0} \mathbf{1}_{(\delta_N < x_{\sigma_{k_0+1}} < \dots < x_{\sigma_{k_0+K_0}}; x_{\sigma_{k_0+j}} \geq \delta_N + (1-\delta_N) \frac{j-1}{K_0})}\right) \\ &\geq \mathbf{P}_0\left(\sum_{|\omega|=k_0} \mathbf{1}_{(x_{\omega_1} < \dots < x_{\omega} < \delta_N)} \geq N^{3/2-\varepsilon}\right) (1 - \mathbf{P}_0(Z_{N, K_0, \delta_N} = 0))^{N^{3/2-\varepsilon}} \\ &= \mathbf{P}_0(Z_{N, k_0}(\delta_N) \geq N^{3/2-\varepsilon}) (1 - (1 - \mathbf{P}_0(Z_{N, K_0, \delta_N} \geq 1)))^{N^{3/2-\varepsilon}} \end{aligned}$$

Or, de même qu'en (4.21), on a $\lambda > 0$ tel que :

$$\mathbf{P}_0(Z_{N, K_0, \delta_N} \geq 1) \geq \lambda N^{-3/2}$$

Et avec l'inégalité de Paley-Zygmund, de la même manière que pour (4.24), on trouve :

$$\mathbf{P}_0(Z_{N, k_0, 1-\delta_N} \geq N^{3/2-\varepsilon}) \geq \frac{1}{C(\gamma)}$$

Or, $Z_{N, k_0}(\delta_N) \geq Z_{N, k_0, 1-\delta_N}$ donc $\mathbf{P}_0(Z_{N, k_0}(\delta_N) \geq N^{3/2-\varepsilon}) \geq \mathbf{P}_0(Z_{N, k_0, 1-\delta_N} \geq N^{3/2-\varepsilon})$

On obtient donc finalement $\tau > 0$:

$$\mathbf{P}_0(Z_{N, eN-3/2 \ln(N)} \geq 1) \geq \frac{1}{C(\gamma)} (1 - (1 - \lambda N^{-3/2})^{N^{3/2-\varepsilon}}) \geq \tau N^{-\varepsilon} \quad (4.26)$$

ce qui permet de conclure.

4.4 Arbre à loi de reproduction binomiale

On présente ici la preuve du résultat intermédiaire permettant d'obtenir finalement l'estimation de $M_N^{(x)}$ donnée en (3.9) en adaptant les preuves données précédemment. Ces adaptations ne seront par contre pas explicitées ici.

On s'intéresse donc au lemme 3.3 : on a une famille $(X_n)_{n \geq 1}$ de variables aléatoires discrètes positives identiquement distribuées d'espérance finie (typiquement de la forme $\mathbf{1}_{(\sigma \text{ est accessible})}$ ou $\mathbf{1}_{\sigma \in \mathcal{A}_L(k)}$) et une variable aléatoire discrète I positive indépendante des précédentes ayant aussi une espérance finie.

On veut alors montrer $\mathbf{E}(\sum_{k=1}^I X_k) = \mathbf{E}(I) \times \mathbf{E}(X_1)$

Démonstration. Par théorème de Fubini, on peut inverser série et espérance puisqu'on travaille dans un cadre positif :

$$\begin{aligned} \mathbf{E}\left(\sum_{k=1}^I X_k\right) &= \mathbf{E}\left(\sum_{k=1}^{+\infty} \mathbf{1}_{I \leq k} X_k\right) \\ &= \sum_{k=1}^{+\infty} \mathbf{E}(\mathbf{1}_{I \leq k}) \mathbf{E}(X_k) \\ &= \mathbf{E}(X_k) \mathbf{E}\left(\sum_{k=1}^{+\infty} \mathbf{1}_{I \leq k}\right) \\ &= \mathbf{E}(X_k) \mathbf{E}(I) \end{aligned}$$

□

Chapitre 5

Discrétisation et modélisation de l'arbre N -aire

Dans ce chapitre, on se propose de modéliser légèrement le modèle en discrétisant les valeurs de survivabilité afin de pouvoir effectuer quelques simulations informatiques.

5.1 Discrétisation des valeurs de survivabilité

Bien que robuste et abordable, l'étude dans le cadre asymptotique $N \rightarrow \infty$ de l'arbre N -aire présente une faiblesse majeure : une très faible capacité de modélisation. En effet, pour un tel problème, on souhaiterait pouvoir observer les comportements asymptotiques sur des arbres aléatoires manipulables. Mais dans le cadre proposé, deux aspects rendent le calcul explicite d'un tel arbre impossible :

- L'arité très élevée de l'arbre conduit à considérer un nombre exponentiel de noeuds. Même en ne conservant que les noeuds "intéressant" (c'est-à-dire issus d'un chemin croissant dans l'arbre), le résultat (3.1) indique que $Z_{N,N}$ est de l'ordre de e^N .
- Le choix d'une loi uniforme dans $[0; 1]$ pour les valeurs de survivabilité associées à chaque noeud conduirait au calcul (et au stockage), d'une quantité tout aussi importante de nombre réels, ce qui est impossible en pratique.

Une solution possible est alors de discrétiser le problème et de se ramener à des valeurs de survivabilité limitées, que l'on choisira entières et de manière uniforme dans $\{1; \dots; h\}$. Cela permet de régler les deux problèmes soulevés puisque d'une part, les nombres aléatoires n'ont plus besoin d'être stockés tels quels, et d'autre part, les noeuds n'ont pas besoin d'être considérés individuellement, et seule l'information de leur valeur de survivabilité associée est importante. Les détails et limites d'une telle implémentation de l'arbre seront discutés dans la section suivante.

Mais une question demeure dans ce nouveau modèle. En effet, la variable aléatoire $Z_{N,k}$ dénombreait le nombre de noeuds σ accessibles à la profondeur k , c'est à dire les noeuds pour lesquels $x_\emptyset < x_{\sigma_1} < \dots < x_\sigma$. Mais lorsque les variables aléatoires x_{σ_i} suivent une loi uniforme dans $[0; 1]$, la probabilité que deux d'entre elles soient égales est nulle. Ce ne sera pas le cas dans notre modèle approché. Ainsi, nous serons amenés à considérer deux cas. Dans le premier, un noeud σ tel que $x_\emptyset \leq x_{\sigma_1} \leq \dots \leq x_\sigma$ sera considéré comme accessible, tandis que dans le deuxième cas, on exigera que les inégalités soient strictes pour que le noeud soit considéré comme accessible. Pour un arbre aléatoire donné, on est donc amené à considérer les deux variables aléatoires suivantes :

$$Z_{N,k}^{inf} = \sum_{|\sigma|=k} \mathbf{1}_{x_\emptyset < x_{\sigma_1} < \dots < x_\sigma} \text{ et } Z_{N,k}^{sup} = \sum_{|\sigma|=k} \mathbf{1}_{x_\emptyset \leq x_{\sigma_1} \leq \dots \leq x_\sigma}$$

Ces deux variables aléatoires sont complémentaires dans une approximation de $Z_{N,k}$. En effet, si l'on fixe un arbre N -aire aléatoire en prenant les valeurs de survivabilité dans $[0; 1]$ puis qu'on discrétise ces valeurs dans $\{0; \frac{1}{h-1}; \frac{2}{h-1}; \dots, 1\}$ (ce qui se ramène donc à prendre des valeurs dans $\{1; \dots; h\}$), on obtient l'encadrement suivant :

$$Z_{N,k}^{inf} \leq Z_{N,k} \leq Z_{N,k}^{sup}$$

Un dernière question subsiste : quelle valeur de h choisir dans le cadre d'une étude pour N grand. Il apparaît clairement que h doit être dépendant de N afin de garantir d'une part une précision convenable dans cette approximation, et d'autre part un temps de calcul qui reste raisonnable. Dans ce but, on cherche une valeur de h telle que, pour un noeud donné, ses fils prennent tous des valeurs différentes dans $\{1; \dots; h\}$, ce qui permet d'obtenir un comportement plus similaire à celui observé dans le cas continu (où les fils ont presque sûrement des valeurs différentes). En notant \mathcal{D} l'événement : "les N fils d'un noeud ont des valeurs différentes dans $\{1; \dots; h\}$ ", il vient

$$\mathbf{P}(\mathcal{D}) = \binom{h}{N} \frac{N!}{h^N} = \frac{h!}{(h-N)!h^N}$$

En prenant $N \rightarrow \infty$, $h \rightarrow \infty$ et $N = o(h)$ (ce qui paraît raisonnable dans le cadre souhaité), il vient :

$$\mathbf{P}(\mathcal{D}) \sim \left(\frac{h}{h-N}\right)^{h-N} e^{-N} \sqrt{\frac{h}{h-N}} \sim e^{-N^2/2h} \sqrt{\frac{h}{h-N}}$$

Prendre h d'ordre N^2 semble alors cohérent. De plus, une telle valeur peut permettre un calcul réalisable en pratique.

Dans un premier temps, on peut alors étudier le comportement au premier ordre des deux variables aléatoires $Z_{N,k}^{inf}$ et $Z_{N,k}^{sup}$. En effet, on a :

$$\mathbf{E}_1(Z_{N,k}^{sup}) = \mathbf{E}(Z_{N,k}^{sup} | x_\emptyset = 1) = N^k \mathbf{P}(1 \leq x_{\sigma_1} \leq \dots \leq x_{\sigma_k} \leq h) = \frac{N^k}{h^k} \binom{k+h-1}{k}$$

En effet, calculer $\mathbf{P}(1 \leq x_{\sigma_1} \leq \dots \leq x_{\sigma_k} \leq h)$ revient à dénombrer les fonctions croissantes de $\{1; \dots; k\}$ dans $\{1; \dots; h\}$, qui sont au nombre de $\binom{k+h-1}{k}$.

Or, on a l'estimée asymptotique suivante :

Lemme 5.1. Pour $k = o(N)$, et $k \rightarrow \infty$, on a :

$$\binom{n}{k} \sim \left(\frac{ne}{k}\right)^k \frac{1}{\sqrt{2\pi k}} e^{-\frac{k^2}{2n}(1+o(1))} \quad (5.1)$$

Démonstration. D'après la formule de Stirling, dans le cadre où $k \rightarrow \infty$ et $N \rightarrow \infty$, on a :

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \sim \sqrt{\frac{n}{2\pi k(n-k)}} \frac{n^n}{k^k (n-k)^{n-k}}$$

Puisque $k = o(n)$, on a que $\sqrt{\frac{n}{2\pi k(n-k)}} \sim \frac{1}{\sqrt{2\pi k}}$. De plus,

$$\begin{aligned} \left(\frac{n}{n-k}\right)^{n-k} &= \exp\left((n-k)\left(-\ln\left(\frac{n-k}{n}\right)\right)\right) \\ &= \exp\left((n-k)\left(-\ln\left(1-\frac{k}{n}\right)\right)\right) \\ &= \exp\left((n-k)\left(\frac{k}{n} + \frac{k^2}{2n^2} + o\left(\frac{k^2}{n^2}\right)\right)\right) \\ &= \exp\left(k - \frac{k^2}{2n} + o\left(\frac{k^2}{n}\right)\right) \\ &= e^k e^{-\frac{k^2}{2n}(1+o(1))} \end{aligned}$$

On peut ainsi conclure à l'estimée souhaitée. \square

Dès lors, en prenant $h = N^2$, pour $k = o(N^2)$ (ce qui est le cas pour les valeurs de k où la population survit) mais dépendant de N de sorte que $k \rightarrow \infty$ lorsque $N \rightarrow \infty$, on obtient :

$$\mathbf{E}_1(Z_{N,k}^{sup}) = \frac{N^k}{N^{2k}} \binom{k+N^2-1}{k} \sim_{N \rightarrow \infty} \left(\frac{(k+N^2-1)e}{Nk}\right)^k \frac{1}{\sqrt{2\pi k}} e^{-\frac{k^2}{2(k+N^2-1)}(1+o(1))} \quad (5.2)$$

Pour $Z_{N,k}^{inf}$, on trouve :

$$\mathbf{E}_1(Z_{N,k}^{inf}) = N^k \mathbf{P}(2 \leq x_{\sigma_1} < \dots < x_{\sigma_k} \leq h) = \frac{N^k}{(h-1)^k} \binom{h-1}{k}$$

Et donc, pour $h = N^2$ avec les mêmes conditions sur k , on obtient :

$$\mathbf{E}_1(Z_{N,k}^{inf}) = \frac{N^k}{(N^2 - 1)^k} \binom{N^2 - 1}{k} \sim_{N \rightarrow \infty} \left(\frac{eN}{k}\right)^k \frac{1}{\sqrt{2\pi k}} e^{-\frac{k^2}{2(N^2-1)}(1+o(1))} \quad (5.3)$$

Dans les deux cas, on observe qu'en prenant k de la forme αN , on retrouve une valeur critique pour l'espérance de $\alpha = e$, comme dans le cadre du modèle initial.

5.2 Algorithme de modélisation

Discutons ici des détails pratiques de l'implémentation de l'arbre discrétisé pour le calcul de $Z_{N,k}$.

En pratique, afin de modéliser informatiquement la structure d'arbre aléatoire discrétisé, on utilise un tableau de dimensions $k \times h$, qui va permettre de stocker, pour chaque génération $1 \leq g \leq k$ et pour chaque valeur $1 \leq i \leq h$ le nombre de noeuds ayant la valeur i à la génération g et étant issus d'une lignée croissante depuis la racine.

L'algorithme naïf qui prend comme argument N et h est alors le suivant :

1. On pose A le tableau qui aura pour $g^{\text{ème}}$ ligne la liste des noeuds de profondeur g dont le chemin à la racine est croissant.
2. On initialise la première ligne de A à la liste correspondant à la racine : $[1, 0, 0, 0, \dots, 0]$ de longueur h .
3. On répète les instructions suivantes tant que la dernière ligne de A n'est pas uniquement composée de 0 :
 - On pose T_0 une liste de h 0, qui va correspondre aux futurs fils des feuilles de A .
 - (\star) Pour chaque feuille x de A on calcule N lois uniformes entre 0 et h . Pour chaque élément aléatoire y (strictement) plus grand que x ainsi tiré, on ajoute +1 à la $y^{\text{ème}}$ case de T_0 .
 - Puis on ajoute T_0 à la fin de A ce qui correspond bien aux feuilles de A dont le chemin depuis la racine est (strictement) croissant.

Un problème demeure lorsque l'on effectue ainsi une modélisation naïve de chaque noeud. En effet, chacun possède N enfants. Ainsi, à la génération k il faut donc effectuer N^k loi uniforme sur $[0; 1]$ dans le pire des cas. La complexité de cet algorithme devient exponentielle ce qui complique grandement une quelconque exploitation des résultats, et rend inutile la démarche de discrétisation.

Des améliorations existent. On peut penser à utiliser une loi binomiale sur tous les noeuds qui sont dans la même case. Mais en terme de complexité, calculer une loi binomiale de paramètres n et p revient à calculer n lois de Bernoulli, ce qui reste insatisfaisant. On peut cependant approcher la loi binomiale par une loi normale, ce qui s'avère plus efficace car son calcul s'effectue en temps constant.

On modifie alors la ligne (\star) de la manière suivante :

Pour chaque entier k entre 0 et h :

- On appelle s le nombre de feuille de A qui sont dans la case k .
- On calcule pour chaque entier entre k et h une loi normale d'espérance Ns/h et de variance $Ns(h-1)/(h^2)$ ce qui correspond à une loi binomiale de paramètre Ns et $1/h$ et permet d'avoir une approximation du nombre de noeud à ajouter dans chacune des cases de T_0 d'indice plus grand que k .

Concernant la complexité de notre algorithme final, on remarque l'on calcule h^2 lois normales dans le pire des cas à chaque nouvelle ligne. La complexité est donc d'ordre N^5 (puisque le nombre de générations à considérer est d'ordre N) ce qui nous permet de faire des modélisation jusqu'à $N = 50$ alors que l'on était limité à $N = 10$ pour l'algorithme naïf.

Il est important de noter que lors de la modélisation d'un arbre, on calcule Z^{inf} et Z^{sup} en même temps, c'est-à-dire que l'on utilise les mêmes tirages aléatoires. Cela revient à considérer un arbre commun et on à compter sur ce même arbre les chemins strictement croissants et ceux largement croissants. C'est ce qui devrait nous permettre d'encadrer la courbe de $\theta(\alpha)$ donnée en (3.1).

5.3 Observations

On utilise l'algorithme pour différentes valeurs de N en prenant $h = N^2$ dans l'objectif d'observer les similitudes entre notre modélisation et les résultats précédemment démontrés. Il est important de noter que pour pouvoir calculer des arbres N -aire pour $N \geq 5$, on a utilisé des lois normales ce qui peut engendrer quelques écarts avec les résultats théoriques. Cependant, pour $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, on a que l'approximation de la loi binomiale de paramètres n et p par une loi normale de paramètres np et $\sqrt{np(1-p)}$ est très satisfaisante. Cela montre que l'approximation effectuée qui, dans les faits, n'est utilisée que dans ces conditions précises, ne nous éloigne pas trop de la modélisation initiale mais augmente grandement notre puissance de calcul. Tout d'abord, suite à ces considérations, on peut donc s'intéresser à l'efficacité effective de notre algorithme. Pour cela, on peut observer la courbe ci-dessous indiquant le temps d'exécution :

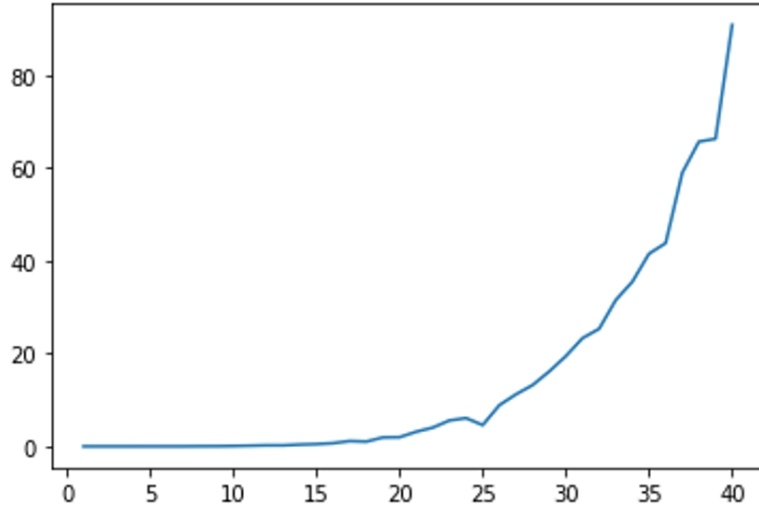


FIGURE 5.1 – Temps de calcul de l'algorithme en secondes en fonction du paramètre N .

On peut alors observer qu'il est réaliste de faire des modélisations jusqu'à $N = 40$ (le calcul s'effectue alors en 1min30). Pour des valeurs de N supérieures, d'après la convexité de la courbe (on rappelle que la complexité est d'ordre N^5) on peut imaginer qu'un arbre d'arité $N = 50$ se calcule en une dizaine de minutes mais qu'il est inutile de chercher à dépasser cette valeur. Cependant, même avec ces valeurs relativement faibles de N , on parvient à observer des résultats intéressants.

On s'attend en premier lieu à ce que les courbes de $Z_{N,k}^{inf}$ et $Z_{N,k}^{sup}$ viennent encadrer celle de $\theta(\alpha)$, qui correspond asymptotiquement au comportement de $Z_{N,k}$. On observe un comportement légèrement différent :

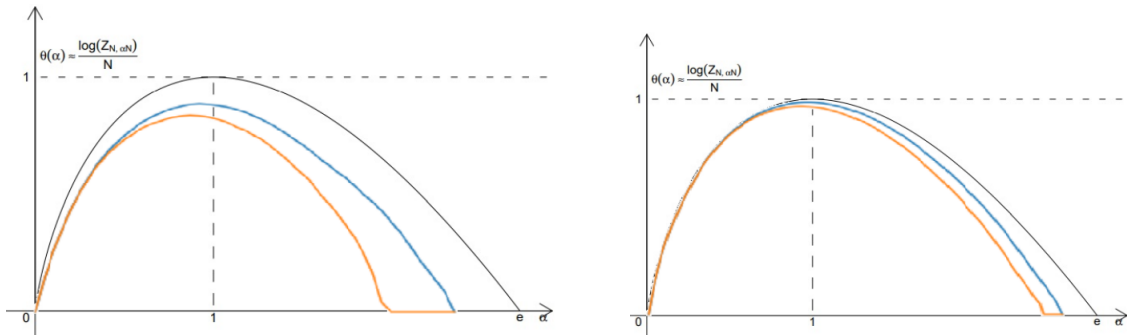


FIGURE 5.2 – $\frac{\ln(Z_{N, \alpha N}^{sup})}{N}$ (en bleu) et $\frac{\ln(Z_{N, \alpha N}^{inf})}{N}$ (en orange) en fonction de α pour deux réalisations d'arbres aléatoires discrets pour $N = 20$ et $N = 50$

En effet, les valeurs de N considérées sont trop faibles, de sorte que même si $Z_{N,k}^{sup}$ est une surestimation, pour $N \leq 50$, $\frac{\ln(Z_{N,k}^{sup})}{N}$ reste inférieur à $\theta(\alpha)$. Il faut donc trouver une courbe théorique dépendant de N à laquelle confronter nos résultats expérimentaux. On va donc utiliser la quantité suivante :

$$\frac{\ln(\mathbf{E}_0(Z_{N,\alpha N}))}{N} = \ln\left(\frac{N^{\alpha N}}{\alpha N!}\right) \frac{1}{N} = \theta(\alpha) - \frac{\ln(2\pi\alpha N)}{2N} + o_{N \rightarrow \infty}(1/N)$$

On obtient alors des comportements qui semblent correspondre au modèle théorique, et donc on peut s'attendre à ce qu'une simulation pour N très grand finisse par donner un bon encadrement de la courbe de $\theta(\alpha)$.

Les courbes suivantes présentent donc une comparaison entre les résultats obtenus (en orange pour $Z_{N,\alpha N}^{inf}$ et en bleu pour $Z_{N,\alpha N}^{sup}$) et la courbe théorique de $\frac{\ln(\mathbf{E}_0(Z_{N,\alpha N}))}{N}$ (en vert).

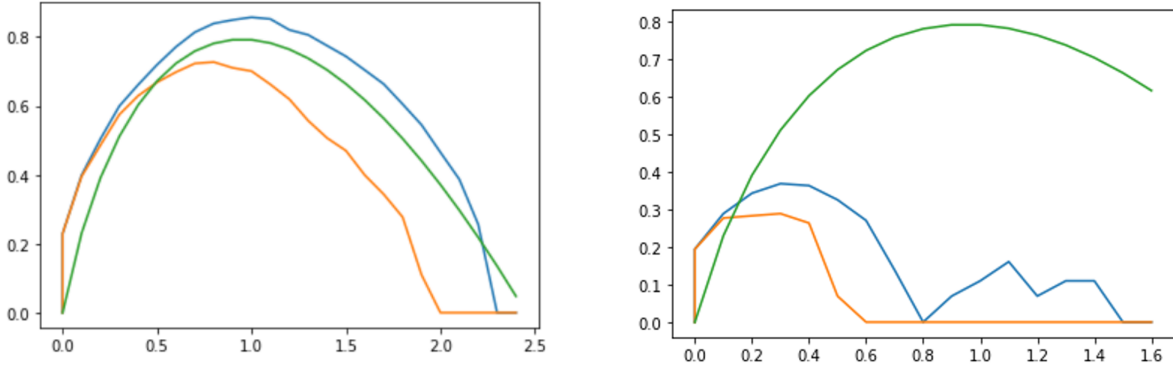


FIGURE 5.3 – $\frac{\ln(Z_{N,\alpha N})}{N}$ en fonction de α pour deux réalisations d'arbres aléatoires discrets avec $N = 10$

Dans la figure précédente, on observe que, bien que dans la majorité des cas l'espérance est bien située entre Z^{inf} et Z^{sup} , il n'est pas improbable que l'on se retrouve dans des cas dégénérés très éloignés du modèle original.

Cependant lorsque l'on augmente N il n'y a plus de cas dégénérés (la convergence de la variable $Z_{N,\alpha N}$ a lieu presque sûrement) et la courbe de l'espérance se retrouve bien encadrée par Z^{inf} et Z^{sup} , comme ci-dessous :

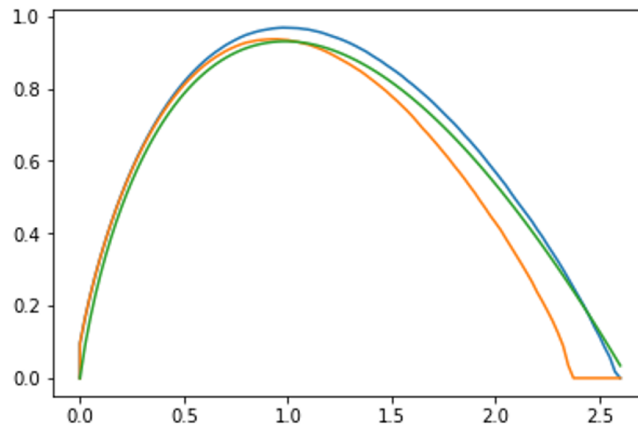


FIGURE 5.4 – Réalisation d'arbre pour $N = 40$

Et finalement, on peut observer que les courbes expérimentales dévient toujours de la courbe théorique pour α proche de e , et ce pour marquer une extinction légèrement prématuré. Cette observation semble faire écho à l'estimation en $eN - 3/2 \log(N)$ pour la génération d'extinction. Malheureusement, les N utilisés étant trop faibles, on ne peut vérifier la valeur numérique $\beta = 3/2$ en pratique.

Bibliographie

- [BBS16] Julien Berestycki, Eric Brunet, and Zhan Shi. The number of accessible paths in the hypercube, 2016.
- [Che14] Xinxin Chen. Increasing paths on n-ary trees, 2014.
- [Kin78] J.F.C Kingman. A simple model for the balance between selection and mutation, 1978.