

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ MOULOUD MAMMERI DE TIZI-OUZOU

FACULTÉ DES SCIENCES

DÉPARTEMENT DE MATHÉMATIQUES

MÉMOIRE DE MASTER

SPÉCIALITÉ : MATHÉMATIQUES

OPTION : PROBABILITÉS-STATISTIQUE

Présenté par

Gherrar Nassima

Ourrad Adélia

Thème:

Processus de Branchement de Galton-Watson et Applications

Devant le jury d'examen composé de :

M. Fellag Hocine;	Professeur ;	UMMTO	Président
M. Hamadouche Djamel;	Professeur;	UMMTO	Rapporteur
M. Boudiba Mohand Arezki ;	M. de conférences A;	UMMTO	Examineur

Soutenu le : 03 /10 /2013

Remerciements

*Nous tenons à exprimer toute notre reconnaissance à M^r **Hamadouche Djamel** pour l'honneur qu'il nous a fait en assurant la direction et le suivi scientifique et technique du présent mémoire. Nous le remercions pour sa grande contribution à l'aboutissement de ce travail, et pour sa disponibilité.*

*Nous remercions vivement M^r **Fellag Hocine** pour l'honneur qu'il nous fait en acceptant de présider le jury de ce mémoire.*

*Nos remerciements chaleureux s'adressent également à M^r **Boudiba Mohand Arezki**, et pour avoir accepté d'examiner ce travail. Ainsi M^{me} **Atil Lynda** pour son aide fidèle.*

Nous remercions tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.

Table des matières

Introduction générale	3
1 Généralités sur les processus de branchement	6
1.1 Processus de branchement	6
1.2 Processus de branchement markovien à temps discret	7
1.2.1 Processus de Galton-Watson	7
1.3 Processus de branchement markoviens à temps continu	9
1.3.1 Processus de branchement de Lévy	11
1.3.2 Processus de branchement de Bellman-Harris	12
1.3.3 Processus de branchement de Crump-Mode-Jagers	16
2 Processus de Galton-Watson	18
2.1 Introduction	18
2.2 Présentation du processus de Galton-Watson	18
2.3 Fonctions génératrices	22
2.4 Probabilité d'extinction	25
2.4.1 Cas sur-critique	30
2.4.2 Cas sous-critique	31
2.4.3 Cas critique	36
2.5 Processus de Galton-Watson avec immigration	37
3 Quelques applications du processus de Galton-Watson	41
3.1 Introduction	41
3.2 Exemple 1	42
3.2.1 Présentation du modèle	42
3.2.2 Etude du phénomène d'extinction	43
3.3 Exemple 2	44
3.3.1 Présentation du modèle	44

<i>TABLE DES MATIÈRES</i>	2
4 Simulations numériques	48
4.1 Résultats et analyse	48
4.1.1 Cas sous-critique et critique	48
4.1.2 Cas sur-critique	52
Conclusion générale	56
Annexe	57
A Code source	57
Bibliographie	60

Introduction générale

Il n'aura échappé à personne que les individus d'une même espèce ne sont pas identiques. Les différences entre les êtres sont expliquées en partie par la variabilité génétique. Une variabilité si forte qu'à de rares exceptions près, chacun peut se considérer comme unique. La génétique des populations a pour but de comprendre d'où provient ce foisonnement de caractères et dans quelle mesure la variabilité génétique est partagée entre les facteurs démographiques, le type de reproduction, la sélection naturelle ou des fluctuations dues au hasard. La complexité de cette étude n'est plus à démontrer et la modélisation de certains phénomènes s'avère être une tâche extrêmement délicate. Les premières et les plus célèbres représentations mathématiques de l'évolution des populations sont les modèles des processus de branchement.

L'origine de ces processus remonte à Bienaymé (1845) qui dans une courte note sur "la loi de la multiplication et la durée des familles", considère un modèle à temps discret où tous les hommes ont la même probabilité d'avoir un nombre de fils donné qui parviennent à l'âge adulte. Bienaymé cherche à calculer la probabilité pour qu'un homme ait encore des descendants qui portent son nom au bout de n générations. Il met en évidence les cas sous-critique, critique ou sur-critique où la moyenne de la loi de reproduction est inférieure, égale ou supérieure à 1 et étudie le comportement d'une lignée quand le nombre de générations tend vers l'infini.

Malheureusement son travail passe complètement inaperçu à l'époque et tombe dans l'oubli. En 1874, Galton et Watson publient "On the probability of extinction of families". Leur but est également de calculer la probabilité d'extinction d'un nom de famille. Le modèle sous-jacent est identique à celui de Bienaymé, Galton et Watson ont l'idée d'introduire la fonction génératrice de la loi de reproduction à la génération n et d'utiliser des relations de récurrence entre les générations n et $n+1$. Galton est le jeune cousin de Darwin et stimulé par la publication de l'origine des espèces, il souhaite comprendre et modéliser la transmission de l'hérédité. C'est dans ce cadre qu'il s'intéresse à l'extinction des familles et l'on peut noter l'analogie entre l'unique dessin du livre de Darwin et celui d'un processus connu maintenant sous le nom de processus de Galton-Watson et qui représente la lignée issue d'un individu.

Ce processus peut également être repéré par la taille de la population à chaque génération n . Chaque reproduction ou mort entraîne un changement de cette taille. C'est ce point de vue que nous allons développer ensuite. Dans ces exemples historiques, le temps est discret et représente les générations successives ou modélisent des espèces qui se reproduisent de manière périodique (comme les reproductions saisonnières). Dans ce travail, nous allons plus spécifiquement nous intéresser à des espèces qui se reproduisent en des temps aléatoires qui peuvent prendre toute valeur réelle. Ainsi les durée de vie et les temps de reproduction de chaque individu sont aléatoires. Ces modèles ont été introduits plus tardivement par McKendrick en 1914 et surtout par Yule qui en 1924, développe de manière détaillée un modèle mathématique de l'évolution des espèces en temps continu pour rendre compte des observations sur ces distributions d'espèces. Nous renvoyons au livre "Histoire de mathématiques et de populations" de N. Bacaer [6] pour plus de détails sur ces premiers travaux. Depuis cette époque, l'étude des processus de branchement continue à se développer et de nombreuses questions, dont beaucoup liées à des problématiques biologiques, sont posées et non résolues. Pour une vision globale sur ces problématiques, nous renvoyons au livre de P. Haccou, P. Jagers, V.A. Vatutin "Branching processes: variation, growth and extinction of populations" [2].

Notre but est d'étudier, en utilisant un modèle mathématique, l'évolution temporelle, ou "dynamique", d'un phénomène. L'existence d'un modèle mathématique permet de pouvoir quantifier numériquement certaines propriétés et de pouvoir prédire certains comportements: par exemple, montrer qu'une certaine population va s'éteindre et calculer l'espérance du temps d'extinction.

Ce mémoire comporte quatre chapitres, le premier est consacré aux grandes classes des processus de branchement, markoviens à temps discret dont le processus de Galton-Watson avec immigration et sans immigration, ainsi que les processus de branchement markoviens à temps continu, tels que le processus de Lévy, le processus de Crump-Mode-Jagers et le processus de Bellman-Harris.

Le second aborde la modélisation des évolutions de population et le calcul des probabilités d'extinction à l'aide du processus de Galton-Watson, dont la fonction génératrice constitue l'outil principal.

Le troisième chapitre traite deux exemples d'applications, l'un s'intéresse aux réplifications d'une macro-molécule d'ADN qui consiste en une chaîne de ϱ nucléotides. En une unité de temps, cette chaîne est répliquée, chaque nucléotide étant copié de façon correcte avec probabilité p et indépendamment des autres nucléotides. A l'issue de la réplification, la molécule est détruite avec probabilité q ou bien donne naissance à deux molécules avec probabilité $1 - q$. On s'intéresse alors à la probabilité de disparition de la population de macro-molécules correctes.

L'autre modélise l'évolution d'une population de parasites dans une cellule, nous considérons le cas où la répartition des parasites est binomiale de paramètres a et p .

Alors quand est ce que l'organisme guérit? de manière équivalente la population de parasites va-t-elle s'éteindre?

Le dernier chapitre vérifie le bien-fondé de la théorie du second exemple donné précédemment à travers des simulations numériques avec le logiciel R. On termine le travail par une conclusion générale et perspectives.

Chapitre 1

Généralités sur les processus de branchement

1.1 Processus de branchement

Les processus de branchement sont des modèles introduits pour étudier le développement d'une population, dans laquelle les individus se reproduisent indépendamment les uns des autres. Ces modèles sont particulièrement utilisés en biologie (étude de la croissance d'une colonie de bactéries,...) et en physique nucléaire, mais trouvent leur origine dans l'étude, au 19^{ème} siècle, des probabilités d'extinction des noms de familles illustres en Grande Bretagne (Francis Galton et Henry Watson, 1874). Leur problème était le suivant:

Si un homme a une probabilité p_0 de n'avoir aucun fils, p_1 d'avoir un fils, p_2 d'en avoir deux, etc; si chacun de ses fils éventuels est dans le même cas, et ainsi de suite; quelle est la probabilité pour qu'à terme cette branche de la famille s'éteigne? Plus généralement, comment connaître la probabilité pour qu'il y ait exactement k individus à la génération n ?

Nous allons cependant essayer de présenter les grandes classes des processus de branchement markovien à temps discret dont le processus de Galton-Watson avec immigration et sans immigration, ainsi que les processus de branchement markoviens à temps continu, tels que le processus de Lévy, le processus de Crump-Mode-Jagers et le processus de Bellman-Harris.

1.2 Processus de branchement markovien à temps discret

1.2.1 Processus de Galton-Watson

Galton est le jeune cousin de Darwin et stimulé par la publication de l'origine des espèces, il souhaite comprendre et modéliser la transmission de l'hérédité. C'est dans ce cadre qu'il s'intéresse à l'extinction des familles et l'on peut noter l'analogie entre l'unique dessin du livre de Darwin fig(1.1) et celui d'un processus connu maintenant sous le nom de processus de Galton-Watson et qui représente la lignée issue d'un individu, Fig(1.2).

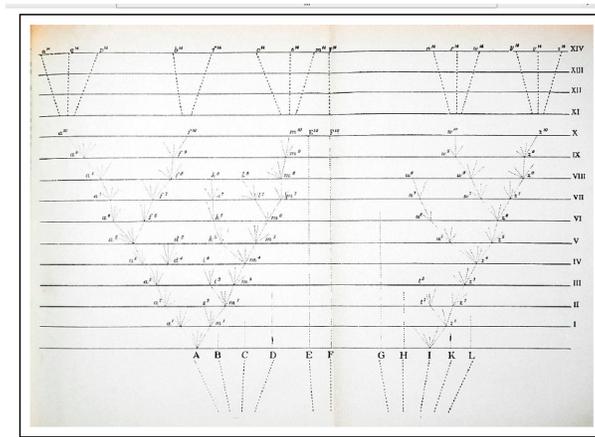


FIG. 1.1 – *Dessin de Darwin dans "Origine des espèce"*

Sur la figure (1.2), on voit que le processus de population peut également être repéré par la taille de la population à chaque génération n . Chaque reproduction ou mort entraîne un changement de cette taille. C'est ce point de vue que nous allons développer ensuite.

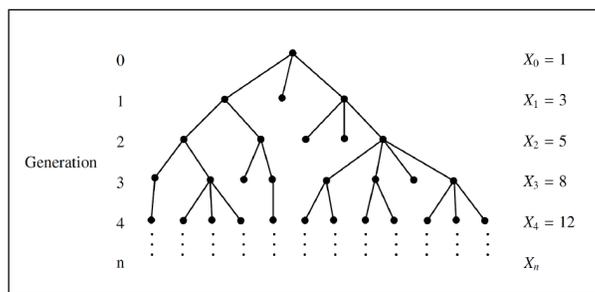


FIG. 1.2 – *Arbre généalogique d'un processus de Galton-Watson*

Dans les modèles en temps discret, le temps est en général exprimé en nombre de générations, mais peut aussi dans certains cas correspondre simplement à une discrétisation simple du temps continu. La variable Z_n est l'effectif de la population à la génération n , elle est donc à valeurs dans \mathbb{N} . Nous étudierons le modèle général suivant:

Si nous supposons qu'à un instant n , la population est de taille $i \in \mathbb{N}(Z_n = i)$, alors chaque individu vivant à la génération n , ou bien

- Meurt lorsque l'on passe à la génération $n + 1$,
- Donne naissance à un nombre aléatoire de descendants, indépendamment des autres individus à la génération $n+1$.

On suppose de plus que les variables "nombre de descendants" ont une loi qui ne dépend pas de l'individu de la n -ième génération dont les descendants sont issus, mais qui peut éventuellement dépendre de i . On la note $\mathcal{P}^i = (p_k^i, i, k \in \mathbb{N})$.

Ainsi, p_k^i est la probabilité qu'un individu issu d'une population de taille i donne naissance, à la génération suivante à k individus.

- En outre, il y a possibilité d'immigration d'un nombre aléatoire d'individus, indépendamment du processus de naissance et de mort, dont la loi $\eta^i = (\eta_j^i, j, i \in \mathbb{N})$ peut aussi dépendre de i .

Ainsi η_j^i désigne la probabilité qu'un nombre j de migrants rejoigne une population de taille i à la génération suivante. Lorsque les migrants sont arrivés, ils se reproduisent ou meurent comme les autres individus de la population.

Construction de la chaîne:

On se donne sur un espace (Ω, \mathcal{A}) une famille de variables aléatoires à valeurs entières

$$(Z_0, (X_{n,i,k})_{n,i,k \geq 0}, (Y_{n,i})_{n,i \geq 0})$$

- La variable aléatoire Z_0 modélise le nombre d'individus au temps 0.
- $X_{n,i,k}$ le nombre d'enfants du k -ième individu de la n -ième génération, si le nombre total d'individus de cette génération est i .
- $Y_{n,i}$ modélise le nombre de migrants arrivant à la $(n + 1)$ -ième génération, sous la même hypothèse $Z_n = i$.

Définition 1.2.1. Pour chaque probabilité μ sur \mathbb{N} , on note \mathbb{P}_μ l'unique probabilité sur l'espace (Ω, \mathcal{A}) sous laquelle ces variables sont indépendantes et telles que

- Z_0 et de loi μ .
- Pour chaque n, i, k , la variable aléatoire $X_{n,i,k}$ est de loi \mathcal{P}^i .
- Pour chaque n, i la variable aléatoire $Y_{n,i}$ est de loi η^i .

Partant de Z_0 , on définit donc les variables aléatoires Z_n par récurrence :

$$Z_{n+1} = \sum_{k=1}^{Z_n} X_{n,i,k} + Y_{n,i} \quad \text{si } Z_n = i, \text{ pour } i \in \mathbb{N}.$$

On note \mathbb{P}_i la probabilité \mathbb{P}_μ lorsque μ est la masse de Dirac en i , cela veut dire que l'on part de $Z_0 = i$ individus.

Définition 1.2.2. Lorsque les probabilités \mathcal{P}^i et η^i sont indépendantes de i , on dit que la dynamique de la population est densité-indépendante. Dans le cas contraire, on dira que la dynamique de la population est densité-dépendante.

On dit qu'il n'y a pas d'immigration lorsque $\eta_0^i = 1$, c'est à dire que η^i est la masse de Dirac en 0 pour tout i . La population considérée est alors dite une population isolée.

On étudiera tout d'abord des modèles de population densité-indépendante et sans immigration. Dans ce cas, la loi de reproduction ne dépend pas du nombre d'individus présents. Ce modèle ne peut être réaliste que dans des situations de ressources infinies. En effet, dans ce cas, les individus ne sont pas en compétition pour survivre et chacun peut se reproduire librement et indépendamment des autres. En revanche, si les ressources (alimentaires par exemple) sont limitées, alors les individus vont entrer en compétition et le nombre d'individus de la population va avoir un effet sur la loi de reproduction. Ce processus est le modèle qui sera étudié dans le chapitre 2.

1.3 Processus de branchement markoviens à temps continu

Considérons un processus $(Z_t)_{t \geq 0}$ décrivant la dynamique de population suivante :

- Au temps $t = 0$, on a un nombre aléatoire Z_0 d'individus.
- Chaque individu à un temps de vie aléatoire qui suit une loi exponentielle de paramètre a .
- Au bout de ce temps, l'individu se reproduit ou meurt suivant la loi de reproduction $(p_i)_{i \in \mathbb{N}}$. La probabilité de mourir est donc p_0 et on suppose que $p_1 = 0$.

- Les temps de vie et les nombres d'enfants de chaque individu sont indépendants les uns des autres.

Définition 1.1. On appelle processus de branchement en temps continu le processus $(Z_t)_{t \geq 0}$ ainsi défini. Z_t représente le nombre d'individus présents au temps t .

Quand $p_0 + p_2 = 1$, c'est-à-dire si on ne peut avoir qu'au plus un descendant, le processus est appelé **processus de branchement binaire** ou **processus de naissance et de mort linéaire**. Un tel processus modélise par exemple le mécanisme de division cellulaire. Si de plus, $d = 0$ (les individus ne meurent jamais), le processus est appelé **processus de fission binaire** ou **processus de Yule**.

Dans la figure, les branches représentent les lignes de descendance des individus. La figure modélise l'évolution temporelle de ces lignées.

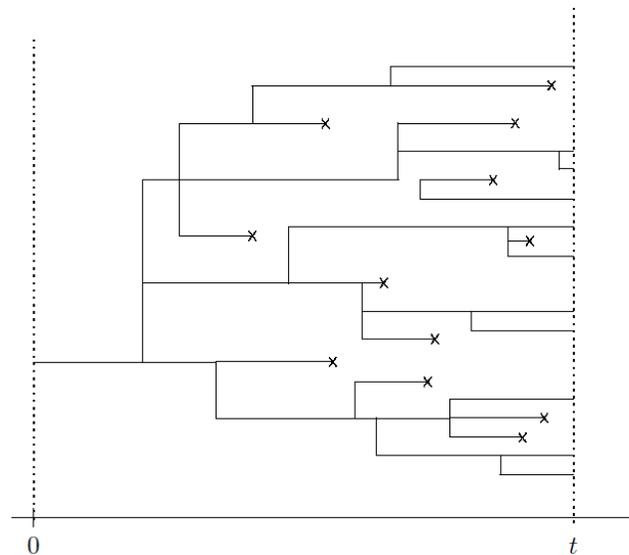


FIG. 1.3 – Un processus de branchement en temps continu

Remarque 1.3.1.

Le fait de modéliser le temps de vie des individus par une loi exponentielle peut être discuté. En effet une loi exponentielle possède la propriété de nonvieillesse qui ne représente pas forcément la réalité.

L'hydre, petit polype d'eau douce de quelques millimètres, en est un bon exemple, ainsi que certaines tortues, certains mollusques ou certains poissons comme l'esturgeon. (Voir [3] sur ce sujet). Dans les autres cas, l'hypothèse de loi exponentielle est liée à la propriété de Markov pour le processus de saut $(Z_t)_{t \geq 0}$. Elle permet donc de faire des calculs en utilisant les équations de Kolmogorov et est considérée comme une bonne approximation par les biologistes.

Signalons toutefois qu'il est possible de généraliser ce modèle en supposant que les temps de vie des individus suivent des distributions plus générales que des lois exponentielles. Le processus est alors appelé processus de Bellman-Harris.

Parmi les processus de branchement markoviens à temps continu, on peut citer le processus de Lévy, le processus de Bellman-Harris et le processus de Crump-Mode-Jagers.

1.3.1 Processus de branchement de Lévy

Un processus de Lévy $X = (X_t, t \geq 0)$ est un processus à accroissements indépendants et stationnaires et dont les trajectoires sont continues à droite et admettent des limites à gauche. Il est dit spectralement positif (ou sans sauts négatifs) lorsqu'il est réel et qu'il n'effectue que des sauts positifs. On note P_x sa loi conditionnelle à $X_0 = x$. L'exposant de Laplace ψ de X appelé aussi mécanisme de branchement est la fonction convexe définie par

$$\mathbb{E}[e^{-\lambda X_t}] = e^{t\psi(\lambda)}, \quad t \geq 0, \quad \lambda > 0$$

qui est caractérisée par la formule de Lévy-Khintchine

$$\psi(\lambda) = \alpha\lambda + \beta\lambda^2 + \int_0^\infty (e^{-\lambda r} - 1 + \lambda r \mathbb{I}_{(r < 1)}) \Lambda(dr).$$

avec $\alpha \in \mathbb{R}$, $\beta \geq 0$ est appelé coefficient gaussien et Λ une mesure de Lévy associée au processus X sur $(0, \infty)$, c'est-à-dire une mesure positive σ -finie telle que

$$\int_0^\infty (1 \wedge r^2) \Lambda(dr) < +\infty.$$

Quand les trajectoires de X sont p.s croissantes, on dit que X est un subordonateur. Dans ce cas, $\psi_\lambda < 0$ pour tout λ strictement positif, et nous préférons définir son exposant de Laplace par $-\psi$. En effet, puisqu'un subordonateur est à variation finie, son exposant de Laplace peut s'écrire :

$$-\psi(\lambda) = d\lambda + \int_0^\infty (1 - e^{-\lambda r}) \Lambda(dr).$$

où $d \geq 0$ est appelé coefficient de dérive. On suppose désormais que X n'est pas un subordonateur. D'abord, puisque ψ est convexe,

$$\lim_{\lambda \rightarrow +\infty} \psi(\lambda) = +\infty.$$

On désigne par η la plus grande racine de ψ

$$\eta := \sup \{ \lambda : \psi(\lambda) = 0 \}.$$

Si $\eta > 0$, ψ a exactement deux racines (0 et η), sinon ψ a pour unique racine $\eta = 0$.

Lorsque $\eta > 0$, X_t tend p.s vers $+\infty$ lorsque $t \rightarrow +\infty$ et son infimum suit la loi exponentielle de paramètre η .

Lorsque $\eta = 0$, soit $\psi'(0^+) > 0$, et X_t tend p.s vers $-\infty$ lorsque $t \rightarrow +\infty$ soit, $\psi'(0^+) < 0$. Ces processus vérifient de nombreuses propriétés. Le problème de double sortie a notamment une solution s'écrivant simplement lorsque les sauts ne sont que positifs en utilisant la fonction d'échelle associée à X , qui est l'unique fonction continue et croissante, $W : [0, \infty] \rightarrow [0, \infty]$, solution de

$$\int_0^\infty e^{-\lambda x} W(x) dx = \frac{1}{\psi(\lambda)}, \quad \lambda > \eta$$

où η est la plus grande racine de ψ , on a alors

$$\mathbb{P}(T_{(-\infty, 0)} < T_{(a, +\infty)} \mid X_0 = x) = \frac{W(a-x)}{W(a)}, \quad 0 < a < x$$

où pour \mathbb{B} un borélien de \mathbb{R} , $T_B = \inf(t \geq 0, X_t \in \mathbb{B})$.

Les processus de Lévy spectralement positifs (et leurs équivalents spectralement négatifs) ont de nombreuses applications : ils peuvent modéliser le capital d'une agence d'assurance, le niveau d'un barrage ou être utilisés en théorie des files d'attente.

1.3.2 Processus de branchement de Bellman-Harris

Les processus de Bellman-Harris, sont la généralisation des processus markoviens au cas où la durée de vie des individus suit une loi quelconque. Ici, nous considérons un processus de branchement de type simple ou nous supposons que chaque individu vit pendant une durée aléatoire L , avec fonction de répartition G et à sa mort, produit un nombre aléatoire X_k d'enfants selon la répartition des descendants p_k .

La reproduction de chaque individu est indépendante de sa durée de vie et celles des autres.

Soit $Z(t)$ le nombre d'individus vivant au temps t . Alors $\{Z(t) : t \geq 0\}$ est appelé un processus à temps continu du type de ramification Bellman-Harris qui a pour durée de vie une distribution G et une loi de reproduction p

Le processus de branchement de Galton-Watson peut être considéré comme un cas particulier de Bellman-Harris lorsque la durée de vie $L = 1$. Ce processus n'est en général pas de Markov, à moins que les durées de vie sont des variables aléatoires indépendantes distribuées de façon exponentielle. Dans un tel cas, le processus est appelé un processus de branchement Markovien en temps continu. Un processus de Bellman-Harris est également appelé un processus de branchement dépendant de l'âge à temps continu.

Le processus de Bellman-Harris est appelé un processus sur-critique, critique ou sous-critique selon que $m > 1$, $m = 1$ ou $m < 1$, tel que

$$m = \sum_{k=1}^{\infty} kp_k.$$

Pour la loi G , nous supposons que $G(0) = 0$, c'est à dire qu'il ya une probabilité nulle de mort instantanée. Harris (1963) a montré cela, ainsi que la production individuelle moyenne finie, cela garantit la finitude du processus, c'est à dire

$$\mathbb{P}(Z_t < \infty) = 1, \quad t \in]0, \infty[.$$

Ensuite, nous introduisons un paramètre α qui décrit le taux de croissance de la population dans le cas sur-critique.

Définition 1.3.1. Soit $\alpha \in \mathbb{R}$ le paramètre malthusien pour la descendance moyenne m et la distribution de vie G . tel que

$$m \int_0^{\infty} e^{\alpha x} dG(x) = 1.$$

En raison de la monotonie du côté gauche de l'équation en fonction de α , une telle racine, quand elle existe, elle est unique. En outre, un tel paramètre malthusien existe toujours et est nécessairement positive ou nulle lorsque $m \geq 1$.

Théorème 1.3.1. [7] Soit f la fonction génératrice de la distribution de la descendance et soit

$$F(s,t) = \sum_{k=0}^{\infty} P(Z_t = k | Z_0 = 1) s^k.$$

alors $F(s,t)$ est la solution unique bornée de l'équation intégrale suivante

$$F(s,t) = s[1 - G(t)] + \int_0^t f(F(s,t-x)) dG(x), \quad |s| \leq 1$$

Nous dirons que F est la fonction génératrice du processus déterminé par (f,G) . Soit q la probabilité d'extinction, c'est-à-dire $q = P(Z_t = 0, t \in]0, \infty[)$. Ensuite, le théorème suivant est une généralisation directe de la vitesse de convergence géométrique de la fonction génératrice d'un processus de Galton Watson.

Théorème 1.3.2. [7]

Si $m \neq 1$, $0 < \gamma = f'(q)$, le paramètre malthusien α pour γ et G existe, et $\mu_\alpha = \gamma \int te^{-\alpha t} dG(t) < \infty$, alors

$$\lim_{t \rightarrow \infty} te^{-\alpha t} (q - F(s,t)) \equiv Q(s).$$

existe pour $0 \leq s \leq 1$ par ailleurs,

$$Q(s) = 0 \text{ si et seulement si } m < 1 \text{ et } \sum_{k=1}^{\infty} (k \log k) p_k = \infty.$$

$$Q(s) \neq 0, s \neq q \text{ si et seulement si } m > 1 \text{ et } \sum_{k=1}^{\infty} (k \log k) p_k < \infty.$$

Théorème 1.3.3. [7] (cas sur-critique) Soit $1 < m < \infty$ alors

(a) Si $\sum_{k=1}^{\infty} (k \log k) p_k < \infty$, alors $e^{-\alpha t} Z_t \rightarrow 0$ w.p.1

(b) Si $Z_0 = 1$ et $\sum_{k=1}^{\infty} (k \log k) p_k < \infty$, alors $e^{-\alpha t} Z_t \rightarrow W$ w.p.1

où W est une variable aléatoire positive ou nulle de telle sorte que

(i) $\mathbb{E}[W] = 1$.

(ii) W est absolument continue sur $(0, \infty)$.

(iii) $\mathbb{P}(W = 0) = q = \mathbb{P}(Z_t = 0) \forall t$

Comme dans les processus de branchement de Galton-Watson en temps discret, il existe des constantes de normalisation "Seneta-Heyde" pour les processus de branchement de Bellman Harris en temps continu.

Schuh et Cohn (1982) ont montré par des approches différentes que si $1 < m < \infty$, sans l'hypothèse de $\sum_{k=1}^{\infty} (k \log k) p_k < \infty$, il existe des constantes C_t tels que:

$$\frac{Z(t)}{C_t} \longrightarrow W.$$

quand $t \longrightarrow \infty$ où W est une variable aléatoire continue sur $(0, \infty)$ de telle sorte que $\mathbb{P}(W = 0) = q = \mathbb{P}(Z_t = 0)$ pour un certain t

Ensuite, lorsque $m = 1$, nous avons un analogue de la loi limite exponentielle du processus critique de branchement de Galton-Watson.

Théorème 1.3.4. [7] (*cas critique*) Si $m = 1$, $\sigma^2 = f''(1) < \infty$, $\mu = \int_0^{\infty} tG(t) < \infty$ et $t^2[1 - G(t)] \rightarrow 0$ quand $t \rightarrow \infty$, alors

$$(a) \lim_{t \rightarrow \infty} \mathbb{P} \left(\frac{Z_t}{t} \leq x | Z_t > 0 \right) = 1 - e^{-\frac{2\mu x}{\sigma^2}}, \quad x \geq 0$$

Dans le cas sous-critique, conditionnée par le cas de non-extinction, le processus Z_t converge vers une variable aléatoire, lorsque $t \longrightarrow 1$. Le résultat est indiqué comme suit.

Théorème 1.3.5. [7] (*cas sous critique*) Si $m < 1$ et $\sum_{k=1}^{\infty} (k \log k) p_k < \infty$. Supposons que le paramètre malthusien α existe pour la descendance moyenne m et la distribution de vie G et $\int_0^{\infty} t e^{-\alpha t} dG(t) < \infty$. Alors

$$(a) \mathbb{P}(Z_t > 0) \sim c e^{\alpha t} \text{ pour } c > 0,$$

$$(b) \text{ Pour tous } k \geq 1, \lim_{t \rightarrow \infty} \mathbb{P}(Z_t = k | Z_t > 0) = b_k \text{ existe } \sum_{k=1}^{\infty} b_k \text{ et } \sum_{k=1}^{\infty} < \infty$$

Pour les preuves de ces derniers résultats, voir Athreya and Ney [4].

Distribution selon l'âge

Un aspect important et utile de processus de branchement dépendant de l'âge, c'est le comportement limite de la répartition par âge.

Considérons un processus de Bellman-Harris ramification $\{Z_t : t \geq 0\}$ qui commence au temps 0 avec une personne de l'âge 0. Cette personne vit pour une durée L avec la distribution vie G et à sa mort, produit ξ enfants selon la répartition $\{p_k\}_{k \geq 0}$ indépendamment des autres personnes vivant dans le même temps. Ensuite, chaque individu vit une durée, meurt et produit sa descendance de la même manière et ainsi de suite.

Nous imposons l'hypothèse que $G(0) = 0$. Nous adoptons aussi les notations suivantes: Pour toute histoire de famille ω

1. $Z(t, \omega)$ est le nombre de personnes vivant au temps t .
2. $Z(x, t, \omega)$ est le nombre d'individus vivants au temps t dont l'âge est inférieur ou égal à x .
3. $A(x, t, \omega) = \frac{Z(x, t, \omega)}{Z(t, \omega)}$.
4. α est le paramètre malthusien pour m et G .
5. $A(x) = \frac{\int_0^x e^{-\alpha u} [1-G(u)] du}{\int_0^\infty e^{-\alpha u} [1-G(u)] du}$.
6. $G_x(t) = \frac{G(x+t) - G(x)}{1-G(x)}$.

1.3.3 Processus de branchement de Crump-Mode-Jagers

Définis indépendamment par K. Crump et C. Mode et par P. Jagers, les processus de branchement généraux, ou processus de Crump-Mode-Jagers (CMJ), sont les processus de branchement les plus généraux possibles. Ce sont des processus de branchement en temps continu et à espace d'états discret qui comptent le nombre d'individus vivants au cours du temps.

À chaque individu X de la population sont associés une variable aléatoire λ_X et un processus ponctuel ξ_X sur $[0, +\infty[$ qui ne sont pas forcément indépendants mais tels que les couples $(\lambda_X, \xi_X)_X$ sont indépendants et identiquement distribués. En fait, λ_X est la durée de vie de X et ξ_X son processus de reproduction, c'est-à-dire qu'il décrit les dates de naissance de ses enfants qui vont évoluer indépendamment et suivant la même dynamique.

Le processus de Crump-Mode-Jagers est alors le processus en temps continu et à espace d'états discret qui compte le nombre d'individus vivants au cours du temps.

Ces processus sont très généraux car les temps de vie ne sont pas forcément exponentiels comme dans les processus de branchement markoviens, les naissances n'ont pas nécessairement lieu à la mort des individus comme dans les processus de Bellman-Harris et le nombre d'enfants d'un individu peut même dépendre de son âge puisque les processus de reproduction et les durées de vie ne sont pas supposés indépendants. Toutes ces raisons font que les processus CMJ peuvent être plus intéressants et pertinents en vue d'applications biologiques. À noter que même les processus de Galton-Watson peuvent être considérés comme des processus CMJ en choisissant des durées de vie constantes égales à 1 et des processus de reproduction ayant leurs atomes en 1.

De nombreux problèmes, concernant notamment les probabilités d'extinction, les taux de croissance des populations et d'autres théorèmes limites, ont été étudiés pour les processus CMJ par des techniques utilisant en particulier la théorie du renouvellement et les caractéristiques aléatoires (voir par exemple [8, 9, 10]). Cependant, même si on arrive à prouver l'existence de limites (en plusieurs sens possibles), celles-ci ne sont en général pas calculables explicitement. Récemment, A. Lambert [11] a obtenu de nouvelles propriétés dans le cas particulier des processus CMJ binaires et homogènes : les naissances n'arrivent qu'une par une et à taux constant $b \in (0, \infty)$ durant la vie des individus.

La distribution commune des durées de vie est notée $\Lambda(\cdot)/b$ où Λ est une mesure positive sur $(0, \infty]$ de masse finie b . Elle est appelée mesure de longévité.

Si (X_t) désigne le nombre d'individus vivants au temps t pour ce cas particulier de processus CMJ, le processus (X_t) n'est en général pas markovien sauf si la mesure de longévité

$$\Lambda(\cdot)/b \sim \varepsilon(d) \text{ où } d > 0.$$

Le processus X_t est alors respectivement un processus de naissance et de mort avec taux de naissance b et taux de mort d , ou un processus markovien de naissance pure à taux b , également appelé processus de Yule. On considère donc une généralisation du processus de naissance et de mort markovien au sens où les naissances sont poissonniennes et les durées de vie sont i.i.d. mais pas nécessairement exponentielles.

Remarque 1.3.2

Dans le cas du processus de Crump-Mode-Jagers, si on se restreint au cas sur-critique ($m > 1$) alors la valeur de m est la solution de cette intégrale

$$m = \int_{(0, \infty)} r \Lambda(dr) > 1.$$

Chapitre 2

Processus de Galton-Watson

2.1 Introduction

Le processus de Galton-Watson est le prototype le plus simple de processus de branchement, défini à temps et espace d'états discret. Le modèle de processus de branchement à été motivé par l'étude de la survie des noms des familles des aristocrates anglais (Galton, 1874), et plus précisément par l'étude de la non-extinction des descendants mâles qui seuls transmettaient le nom.

Le modèle est le suivant: N hommes adultes d'une nation qui portent tous des noms de famille différents partent coloniser un pays. On suppose qu'à chaque génération, la proportion d'hommes qui ont j garçons est p_j , $j \in \mathbb{N}$.

De nombreux phénomènes de dynamique de populations peuvent en fait être modélisés par ce type de processus, dont la description est simple. Le modèle décrit aussi bien la croissance d'une population de cellules, que d'une épidémie dans une population, ou la réplication d'une molécule d'ADN en génétique.

2.2 Présentation du processus de Galton-Watson

La variable Z_n est le nombre de noeuds de profondeur n . La chaîne de vie et de mort la plus simple est la chaîne de Galton Watson C'est une chaîne sans immigration, densité-indépendante. On a

$$\mathcal{P}^i = \mathcal{P}, \quad i \in \mathbb{N}$$

et η^i est la masse de Dirac en 0. Par suite, on étudie l'évolution du nombre Z_n d'individus de cette population au cours des générations successives ($n \geq 0$) en supposant que chacun des Z_n individus de la n -ième génération engendre un nombre aléatoire $X_{n,k}$ d'enfants ($1 \leq k \leq Z_n$) de sorte que

$$Z_{n+1} = \sum_{k=1}^{Z_n} X_{n,k}, \quad n \geq 0. \quad (2.2.1)$$

Les variables aléatoires $X_{n,k}$ sont supposées indépendantes entre elles et de même loi. Cette loi, est une probabilité sur \mathbb{N} , définie par $(p_j, j \in \mathbb{N})$ et caractérisée par la fonction génératrice $g(s)$. Dans toute la suite, nous excluons le cas inintéressant où X serait égale à 1 avec probabilité un, c'est-à-dire le cas où le cas $g(s) \equiv s$.

Remarque 2.2.1 (Arbre de Galton-Watson).

La suite $(Z_n)_{n \geq 0}$ issu de $Z_0 = 1$, se représente comme un arbre aléatoire dont la loi de branchement est \mathbb{P} . La variable Z_n est le nombre de noeuds de profondeur n .

L'arbre de Galton-Watson est un objet mathématique aléatoire utilisé dans la théorie des probabilités. C'est un arbre planaire enraciné, chaque nœud a un nombre aléatoire de fils et ce nombre ne dépend ni de la position du nœud dans l'arbre ni du nombre de fils des autres nœuds.

Un formalisme des arbres de Galton-Watson a été introduit, par Jaques Neuveu en 1986, comme représentation de la généalogie des processus de Galton Watson.

La figure 2.1 représente les individus des générations 0 à 3, lorsque $X_0 = 1, X_{0,1} = 2, X_{1,1} = 3, X_{1,2} = 1, X_{2,1} = 2, X_{2,2} = 0, X_{2,3} = 0, X_{2,4} = 3$.

Les lignes représentent les générations Z_n et les nœuds le nombre d'individus $X_{n,k}$.

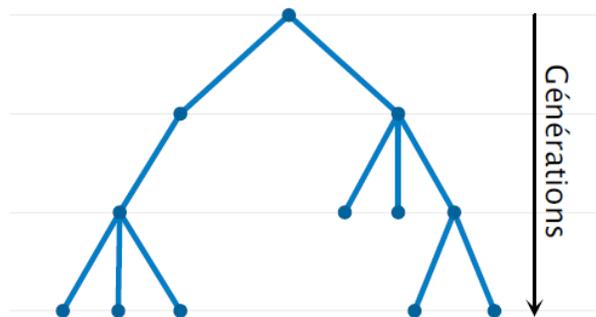


FIG. 2.1 – Arbre généalogique d'un processus de Galton-Watson.

Considérons par exemple (voir Kimmel-Axelrod [2], Delmas-Jourdain [3]) une macromolécule d'ADN qui consiste en une chaîne de ϱ nucléotides. En une unité de temps, cette chaîne est répliquée, chaque nucléotide étant copié de façon correcte avec probabilité p et indépendamment des autres nucléotides. A l'issue de la réplication, la molécule est détruite avec probabilité q ou bien donne naissance à deux molécules avec probabilité $1 - q$. On s'intéresse alors à la probabilité de disparition de la population de macromolécules correctes. Elle est égale à la probabilité d'extinction d'un nom de famille dans le cas où $p_0 = q$ (destruction), $p_1 = (1 - q)(1 - p^e)$ (non destruction mais réplication incorrecte), $p_2 = (1 - q)p^e$ (non destruction et réplication correcte), et $p_k = 0$ pour $k \geq 3$.

Cas particuliers

- Si $p_0 = 0$, $p_1 < 1$, alors $\forall n \in \mathbb{N}$, $Z_{n+1} \geq Z_n$. L'effectif de la population croît.
- Si $p_1 = 1$, $Z_n = Z_0$. L'effectif de la population ne varie pas.
- Si $p_1 < 1$ et $p_0 + p_1 = 1$, on a $Z_{n+1} \leq Z_n$. L'effectif de la population décroît.

Dans la suite, on supposera donc que

$$p_0 > 0 \text{ et } p_0 + p_1 < 1.$$

Proposition 2.2.1. [3] Z_n est une chaîne de markov homogène.

Preuve. $\forall (x, y, z_1, z_2, \dots, z_n) \in \mathbb{N}^{n+1}$,

$$\begin{aligned} & \mathbb{P}[Z_{n+1} = y / Z_n = x, \dots, Z_1 = z_1] \\ &= \mathbb{P}\left(\sum_{k=1}^{Z_n} X_{n,k} = y, Z_n = x, \dots, Z_1 = z_1\right) / \mathbb{P}(Z_n = x, \dots, Z_1 = z_1) \\ &= \mathbb{P}\left(\sum_{k=1}^{Z_n} X_{n,k} = y\right). \end{aligned}$$

Par indépendance des variables $(X_{n,k}, k \geq 1)$ et des variables Z_1, \dots, Z_n qui sont fonctions des $(X_{l,k}, 0 \leq l \leq n - 1, k \geq 1)$.

Ainsi le processus $(Z_n, n \geq 0)$ est une chaîne de Markov homogène de matrice de transition

$$\forall x, y \in \mathbb{N}, P(x, y) = \mathbb{P}\left(\sum_{k=1}^x X_{n,k} = y\right). \quad (2.2.2)$$

Proposition 2.2.2. [6]

$$\begin{aligned} (Z_n) \text{ martingale} &\iff m = 1. \\ (Z_n) \text{ sur - martingale} &\iff m < 1. \\ (Z_n) \text{ sous - martingale} &\iff m > 1. \end{aligned}$$

Preuve. Pour tout $n \geq 1$ et pour tout $(z_1, \dots, z_n) \in \mathbb{N}^n$, soit (\mathcal{F}_n) la filtration naturelle

du processus (Z_n) : (Z_n) est adapté à (\mathcal{F}_n) et positif.

On a

$$\begin{aligned} \mathbb{E}[Z_{n+1}/\mathcal{F}_n] &= \mathbb{E}\left[\sum_{k=1}^{Z_n} X_{n,k}/\mathcal{F}_n\right] \\ &= \sum_{k=1}^{Z_n} \mathbb{E}[X_{n,k}] = mZ_n \end{aligned}$$

car les $X_{n,k}$ sont indépendantes de Z_n .

Donc

$$\mathbb{E}[Z_{n+1}/\mathcal{F}_n] = mZ_n.$$

D'où

$$\begin{aligned} (Z_n) \text{ martingale} &\iff m = 1. \\ (Z_n) \text{ sur - martingale} &\iff m < 1. \\ (Z_n) \text{ sous - martingale} &\iff m > 1. \end{aligned}$$

2.3 Fonctions génératrices

L'outil essentiel d'étude d'un processus de Galton-Watson est la fonction génératrice .

Définition 2.3.1. On note g la fonction génératrice de X , définie par :

$$\forall s \in [0,1], \quad g(s) = \mathbb{E}[s^X] = \sum_{j=0}^{\infty} s^j \mathbb{P}(X = j).$$

Propriétés 2.3.1. [7]

Soit X une variable aléatoire, $g(s)$ sa fonction génératrice, on a alors:

1. $g(1) = 1$ et $g(0) = \mathbb{P}(X = 0)$,
2. $\mathbb{E}(X) = g'(1) = m$,
3. $\mathbb{E}(X(X-1)\dots(X-k+1)) = g^{(k)}(1)$,
4. $\text{Var}(X) = g''(1) + g'(1) + g'(1)^2$.

Lemme 2.3.1. La fonction g est convexe et strictement croissante sur $[0,1]$, elle vérifie

$$g(0) = \mathbb{P}(X = 0) \text{ et } g(1) = 1.$$

Preuve.

Rappelons que la fonction g est définie par

$$g(s) = \sum_{j=0}^{\infty} s^j \mathbb{P}(X = j).$$

où les nombre $\mathbb{P}(X = j)$ sont tous positifs ou nuls, de somme 1.

D'où

$$g(1) = \sum_{j=0}^{\infty} \mathbb{P}(X = j) = 1 \text{ et } g(0) = \mathbb{P}(X = 0) \geq 0 .$$

Chaque terme de la série définissant une fonction strictement croissante, la fonction g est donc elle-même strictement croissante. De plus, g étant une série entière de rayon de convergence inférieure à 1, on peut dériver terme à terme pour obtenir :

$$\forall s \in [0,1], \quad g'(s) = \sum_{j=0}^{+\infty} j p_j s^{j-1}.$$

Ceci montre que la fonction g' est elle-même croissante sur $[0,1[$, et donc la fonction g est convexe sur $[0,1]$.

Il reste uniquement à démontrer que g est continue en 1. Il suffit pour cela de voir que, pour tout entier N et tout réel $s \in [0,1[$, on a

$$\sum_{j=0}^N \mathbb{P}(X = j)s^j \leq g(s) \leq 1.$$

g étant croissante et majorée, elle admet une limite en 1. Passant à la limite en s , il vient

$$\sum_{j=0}^N \mathbb{P}(X = j) \leq \lim_{s \rightarrow 1} g(s) \leq 1.$$

Passant enfin à la limite en N ,

$$\lim_{s \rightarrow 1} g(s) = 1 = g(1).$$

□

Définition 2.3.2. Pour $n \in \mathbb{N}$, on note g_n la fonction génératrice de Z_n ,

$$\forall s \in [0,1], \quad g_n(s) = \mathbb{E}[s^{Z_n}] = \sum_{j=0}^{\infty} s^j \mathbb{P}(Z_n = j).$$

La suite $(g_n)_{n \in \mathbb{N}}$ vérifie cette relation fondamentale .

Théorème 2.3.1. [3]

$$\forall n \in \mathbb{N}, \quad \forall s \in [0,1], \quad g_n(s) = g_{n-1}(g(s)).$$

Preuve.

On utilise le caractère discret de la variable aléatoire Z_{n-1} qui assure que

$$1 = \sum_{j \in \mathbb{N}} \mathbb{I}_{\{Z_{n-1}=j\}} \text{ et son indépendance avec les } X_{n-1,k}, k \geq 1.$$

Ces deux propriétés entraînent que

$\forall n \in \mathbb{N}$ et $\forall s \in [0,1]$,

$$\begin{aligned} g_n(s) &= \mathbb{E}[s^{Z_n}] = \mathbb{E}\left[\sum_{j \geq 0} \mathbb{I}_{\{Z_{n-1}=j\}} s^{\sum_{k=1}^j X_{n-1,k}}\right] \\ &= \sum_{j \geq 0} \mathbb{P}(Z_{n-1} = j) \mathbb{E}[s^{X_{n-1,1}}]^j = \sum_{j \geq 0} g(s)^j \mathbb{P}(Z_{n-1} = j) \\ &= g_{n-1}(g(s)). \end{aligned}$$

Corollaire 2.3.1.

On obtient la formule suivante, importante pour la suite :

$$\forall n \in \mathbb{N}, g_n = g^n (= \underbrace{g \circ \dots \circ g}_{n \text{ fois}}).$$

Lemme 2.3.2. Soit $m = \mathbb{E}(X)$ alors $\mathbb{E}[Z_n] = m^n$.

Preuve.

par le théorème 2.3.1

$$g_n(s) = g(g_{n-1}(s)).$$

En dérivant cette égalité et en l'évaluant en $s = 1$, on aura

$$g'_n(1) = g'(g_{n-1}(1))g'_{n-1}(1). \quad (1)$$

On utilise les propriétés des fonctions génératrices, l'inégalité (1) devient

$$g'_n(1) = g'(1)g'_{n-1}(1) = \mathbb{E}[X]\mathbb{E}[Z_{n-1}] = m\mathbb{E}[Z_{n-1}].$$

Donc

$$\mathbb{E}[Z_n] = g'_n(1) = m\mathbb{E}[Z_{n-1}].$$

D'où, en itérant cette égalité, on obtient :

$$\mathbb{E}[Z_n] = m\mathbb{E}[Z_{n-1}] = m^2\mathbb{E}[Z_{n-2}] = \dots = m^n.$$

On déduit du lemme 2.3.1 que

$$\mathbb{E}[Z_n] \longrightarrow \begin{cases} 0 & \text{si } m < 1, \\ 1 & \text{si } m = 1, \\ +\infty & \text{si } m > 1, \end{cases}$$

lorsque $n \longrightarrow \infty$.

Ce résultat indique que le comportement du processus dépend fortement de la valeur de m .

Remarque 2.3.1.

- En particulier, si nous supposons que $Z_0 = i$, on a

$$G_n(s) = g_n(s)^i,$$

$$\mathbb{E}_i[Z_n] = \mathbb{E}[Z_n/Z_0 = i] = m^n i.$$

- $\mathbb{P}_i(Z_n \neq 0) \leq \mathbb{E}_i[Z_n] = m^n i$.
- Les deux premiers moments (qui peuvent être infinis) sont:

$$m = \sum_{k=0}^{\infty} k p_k \quad , \quad m_2 = \sum_{k=0}^{\infty} k^2 p_k.$$

2.4 Probabilité d'extinction

Une des questions les plus intéressantes à propos des processus de branchement est de savoir si oui ou non il va continuer indéfiniment. On a vu précédemment que l'espérance du nombre d'individus dépendait de l'espérance de la variable aléatoire Z_1 .

En effet, cette dépendance est assez logique si l'on considère le fait que

$m < 1$: signifie que chaque individu donne naissance à (en moyenne) strictement moins de descendants qu'il n'en faut pour combler le vide causé par sa mort.

$m > 1$: signifie que chaque disparition d'un individu entraîne un accroissement de la population.

$m = 1$: est appelé critique car l'espérance des Z_n vaut 1 pour tout n .

l'extinction de la population est certaine si et seulement si $m \leq 1$.

Définition 2.4.1. Soit $\zeta = \inf\{n \geq 0; Z_n = 0\} \in \mathbb{N} \cup \{+\infty\}$. La probabilité d'extinction de la population partant d'un nombre i d'individus, notée θ_i , est définie par:

$$\theta_i = \mathbb{P}_i(\zeta < +\infty), \quad i \in \mathbb{N}^*.$$

Propriétés 2.4.1. *En particulier, si nous supposons que $Z_0 = 1$, on a*

$$\theta_1 = \theta = \mathbb{P}(\lim_{n \rightarrow \infty} Z_n = 0) = \mathbb{P}(\bigcup_{n \geq 0} \{Z_n = 0\}) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0)$$

Preuve.

Remarquons que si $Z_n = 0$ alors $Z_{n+1} = 0$, donc les événements $\{Z_n = 0\}$ croissent avec n . L'événement A = "Extinction de la population" est donc naturellement défini par

$$A = (\bigcup_{n \geq 0} \{Z_n = 0\})$$

Donc

$$\{\lim_{n \rightarrow \infty} Z_n = 0\} = \{\bigcup_{n \geq 0} \{Z_n = 0\}\}$$

et comme \mathbb{P} est de mesure pleine, on a le résultat. □

Propriétés 2.4.2. *θ est le plus petit point fixe de g .*

Preuve.

Remarquons d'abord que $\mathbb{P}(Z_n = 0) = g_n(0)$. Donc $\theta = \lim_{n \rightarrow \infty} g_n(0)$ d'après ce qui précède. De plus, par continuité de g et d'après le corollaire 2.3.1, on a

$$\theta = g(\lim_{n \rightarrow \infty} g_{n-1}(0)) = g(\theta)$$

Il reste à montrer que c'est le plus petit point fixe, en effet si ε est un autre point fixe, comme les fonctions g_n sont convexes et croissantes, on a $g_n(0) \leq g(\varepsilon) = \varepsilon$. On fait tendre n vers l'infini et on obtient $\theta \leq \varepsilon$. □

On pose,

$$s_0 = \inf\{s \in [0,1], g(s) = s\}.$$

On a $s_0 > 0$ car $p_0 > 0$ et $s_0 \leq 1$ car $g(1) = 1$.

Lemme 2.4.1.

1. Si $g'(1) \leq 1$, alors $s_0 = 1$, et $g_n(s)$ croît vers 1 lorsque $n \nearrow \infty$ pour tout $s \in [0,1]$.
2. Si $g'(1) > 1$, l'équation $g(v) = v$ possède une solution unique s_0 dans $[0,1[$ et $g_n(s)$ croît vers s_0 (resp. décroît) lorsque $n \nearrow \infty$, pour tout $s \in [0, s_0]$, (resp. tout $s \in [s_0, 1]$).

Preuve.

L'application $s \rightarrow g(s)$ dans l'intervalle $[0,1]$ est croissante et strictement convexe. De plus $g(1) = 1$. Comme nous avons exclu le cas $g(s) \equiv s$, la courbe g ne coupe pas ou au contraire coupe la diagonale du carré $[0,1]^2$ en un point distinct de $(1,1)$, selon que $g'(1) \leq 1$ ou au contraire que $g'(1) > 1$; ceci se voit bien sur la figure (2.2) Ainsi, selon le cas, l'équation de point fixe $g(v) = v$ n'a pas de solution ou au contraire possède une unique solution dans $[0,1[$, c'est à dire que soit $s_0 = 1$, ou $s_0 < 1$

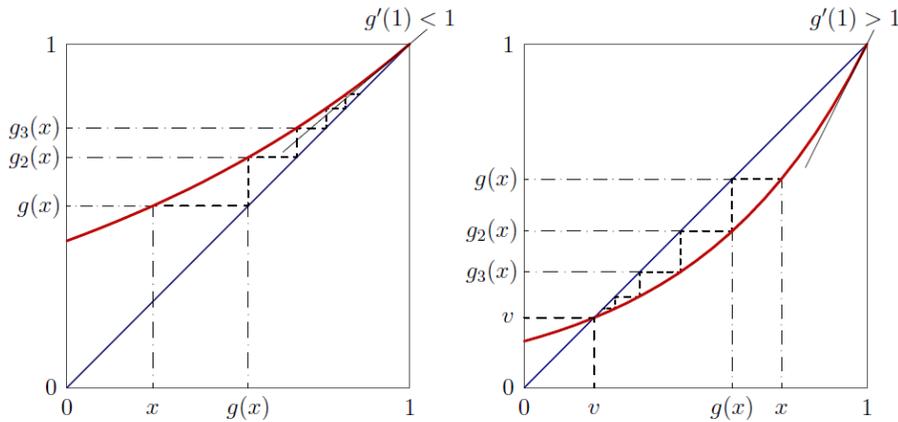


FIG. 2.2 – La fonction g et ses itérés dans les deux cas $m \leq 1$ et $m > 1$ ($v = s_0$).

Lorsque $g'(1) \leq 1$, nous avons $s \leq g(s)$ et donc $g_n(s) \leq g_{n+1}(s)$ (puisque $g_{n+1} = g \circ g_n$) pour tout s . La limite $\lim_{n \rightarrow \infty} g_n(s)$ est inférieure à 1 et solution de $g(u) = u$, elle ne peut alors valoir que 1.

De même, si $g'(1) > 1$, nous avons $s \leq g(s) \leq s_0$ ou $s \geq g(s) \geq s_0$ selon que $s \leq s_0$ ou que $s \geq s_0$; il s'en suit que $g_n(s)$ croît (resp. décroît) avec n selon le cas. La limite $\lim_{n \rightarrow \infty} g_n(s)$ qui est solution de $g(u) = u$ et est strictement inférieure à 1, (du moins si $s \neq 1$), est alors nécessairement égale à s_0 .

Exemples 2.4.1. (Loi de reproduction Poissonienne).

Si $P = Poi(\lambda)$ alors $m = \lambda$ et $g(s) = \exp\{\lambda(s-1)\}$ et donc si $\lambda \leq 1$ alors la population s'éteint presque sûrement tandis que si $\lambda > 1$ alors la probabilité d'extinction θ est l'unique solution sur $]0,1[$ de l'équation $\lambda = \frac{\log(\theta)}{\theta-1}$.

Exemples 2.4.2. (Loi de reproduction géométrique).

Si $\mathbb{P} = \text{Geom}_{\mathbb{N}}(p) = \sum_{n \geq 0} q^n p \delta_n$ alors $m = q/p$ et $g(s) = p/(1 - sq)$ et donc si $p \geq 1/2$ alors la population s'éteint presque sûrement tandis que si $p < 1/2$ alors la probabilité d'extinction s est l'unique solution sur $]0; 1[$ de l'équation $qs^2 - s + p = 0$.

Théorème 2.4.1. [5]

1) Dans les cas sous-critique et critique, ($m \leq 1$), on a $\theta_i = \mathbb{P}_i(\zeta < +\infty) = 1$. La population s'éteint presque sûrement.

2) Dans le cas sous-critique, on a $\mathbb{E}_i[\zeta] < \infty$, tandis que dans le cas critique, on a $\mathbb{E}_i[\zeta] = \infty$.

3) Dans le cas surcritique ($m > 1$), sous \mathbb{P}_i , la suite Z_n converge presque-sûrement vers une limite Z_∞ qui ne prend que les valeurs 0 et $+\infty$ avec les probabilités

$$\theta_i = \mathbb{P}_i(\zeta < +\infty) = \mathbb{P}_i(Z_\infty = 0) = s_0^i, \quad \mathbb{P}_i(Z_\infty = +\infty) = 1 - s_0^i.$$

Preuve.

La démonstration du théorème repose sur le lemme analytique suivant.

Lemme 2.4.2.

- On a $1 - g_n(s) \leq m^n(1 - s)$.
- Si $m = 1$ et $m_2 < \infty$, alors pour tout $s \in [0, 1[$, la suite $1 - g_n(s)$ est équivalente à

$$\frac{2}{n(m_2 - 1)} \text{ quand } n \longrightarrow \infty.$$

Preuve.

Remarquons que $Z_{n+1} = 0$ si $Z_n = 0$, et donc les événements $\{Z_n = 0\}$ croissent avec n . L'événement $A =$ "Extinction de la population" est donc naturellement défini par

$$A = \{\zeta < +\infty\} = \bigcup_n \uparrow \{Z_n = 0\}$$

et sa probabilité donnée par

$$\mathbb{P}(A) = \lim_n \uparrow G_n(0).$$

La démonstration du théorème repose donc sur l'étude la suite $(G_n(0))_n$, avec $G_n(0) = [g_n(0)]^i$. Nous allons utiliser le lemme 2.4.1.

1) On a $\{\zeta < n\} = \{Z_n = 0\}$, et donc

$$\mathbb{P}_i(\zeta \leq n) = [\mathbb{P}_1(\zeta \leq n)]^i = [g_n(0)]^i \quad (2.4.1)$$

converge vers s_0^i quand n tend vers l'infini. Ainsi, cela entraîne que $\mathbb{P}_i(\zeta < \infty) = 1$ si $m \leq 1$ d'ou la première partie du théorème. De plus $\mathbb{P}_i(\zeta < \infty) = s_0^i$ quand $m > 1$.

2) On a aussi

$$\mathbb{E}_i[\zeta] = \sum_{n \geq 0} \mathbb{P}_i(\zeta > n) = \sum_{n \geq 0} (1 - g_n(0)^i).$$

Le lemme 2.4.2 entraîne immédiatement que si $m < 1$, alors $E_1[\zeta] < \infty$, puisque dans ce cas, la série de terme général m^n est convergente. De plus, le temps d'extinction de Z_n sachant que $Z_0 = i$ est le maximum des temps d'extinction des sous populations issues de chaque individu. Ainsi, pour tout $i \in \mathbb{N}^*$, $\mathbb{E}_i[\zeta] < \infty$.

Si $m = 1$ et $m_2 < \infty$, alors par la deuxième partie du lemme, on sait que $\mathbb{E}_1[\zeta] = \infty$ (la série harmonique diverge), ce qui entraîne alors que $\mathbb{E}_i[\zeta] = \infty$ pour tout $i \geq 1$. On a donc montré l'assertion (2) du théorème.

3) Il reste à étudier le comportement asymptotique de Z_n quand $m > 1$.

Lorsque $m > 1$, l'extinction n'est pas certaine puisque $\mathbb{P}(A) = s_0 < 1$. Rappelons que l'état 0 est absorbant et que les autres états sont transients. Donc deux cas seulement sont possibles : ou bien $Z_n = 0$ à partir d'un certain rang (c'est l'extinction), ou bien $Z_n \rightarrow \infty$ (c'est l'explosion). Ces deux événements complémentaires ont pour probabilités respectives s_0^i et $1 - s_0^i$. Cela conclut la preuve du théorème. \square

Remarque 2.4.1.

La formule (2.4.1) donne en fait la fonction de répartition du temps d'extinction ζ , et donc sa loi, sous \mathbb{P}_1 et sous \mathbb{P}_i . On a

$$\mathbb{P}_1(\zeta = n) = \begin{cases} 0 & \text{Si } n = 0; \\ g_n(0) - g_{n-1}(0) & \text{Si } n \in \mathbb{N}^*; \\ 1 - s_0 & \text{Si } n = +\infty. \end{cases}$$

Le théorème précédant nous amène à distinguer 3 cas :

- Cas sous-critique ($m < 1$).
- Cas critique ($m = 1$).
- Cas sur-critique ($m > 1$).

2.4.1 Cas sur-critique

Théorème 2.4.2. [5] (*cas sur-critique*) Si $m > 1$ et $m_2 < \infty$, la suite $\frac{Z_n}{m^n}$ converge p.s. vers une variable aléatoire W qui est presque-sûrement strictement positive sur l'ensemble $\{\zeta = \infty\}$ et telle que $\mathbb{E}_i[W] = i$.

Ce résultat nous dit que sur l'ensemble où il n'y a pas extinction, la croissance de Z_n vers $+\infty$ est exponentielle (en m^n). C'est l'expression probabiliste de la **loi de Malthus**.

Preuve.

Posons $W_n = \frac{Z_n}{m^n}$. Considérons la filtration $(\mathcal{F}_n)_n$ formée de la suite croissante de tribus \mathcal{F}_n , où \mathcal{F}_n est la tribu engendrée par les $(Z_k, k \leq n)$. Comme m et m_2 sont les deux premiers moments des variables $X_{n,k}$ on a, en utilisant (2.2.1)

$$\mathbb{E}_i[Z_{n+1}/\mathcal{F}_n] = mZ_n \text{ et } \mathbb{E}_i[Z_{n+1}^2/\mathcal{F}_n] = m_2Z_n + m^2Z_n(Z_n - 1)$$

et donc

$$\mathbb{E}_i[W_{n+1}/\mathcal{F}_n] = W_n \text{ et } \mathbb{E}_i[W_{n+1}^2/\mathcal{F}_n] \leq W_n^2 + \frac{m_2}{m^{n+2}}W_n$$

Par suite, $\mathbb{E}_i[W_n] = i$ et comme $m > 1$,

$$\mathbb{E}_i[W_{n+1}^2] \leq \mathbb{E}_i[W_n^2] + \frac{im_2}{m^{n+2}} \leq i^2 + \frac{im_2}{m^2} \sum_{j=0}^n m^{-j} \leq i^2 + \frac{im_2}{m(m-1)}$$

On en déduit finalement que W_n est une martingale, dont les moments d'ordre 2 sont uniformément bornés. Elle converge donc \mathbb{P}_i -presque-sûrement et en moyenne vers W .

En particulier, $\mathbb{E}_i[W] = i$. Il reste à montrer que $W > 0$, p.s. sur $\{\zeta = \infty\}$.

On pose $r_i = \mathbb{P}_i(W = 0)$. En vertu de la propriété de branchement, la loi de W sous \mathbb{P}_i est la même que celle de $W^1 + \dots + W^i$, où les W_j sont des variables aléatoires indépendantes de même loi que W sous \mathbb{P}_1 .

Par suite, on a $r_i = (r_1)^i$. Par la propriété de Markov, on a

$$\mathbb{P}_1(W = 0/\mathcal{F}_1) = \mathbb{P}_{Z_1}(W = 0) = r_1^{Z_1}.$$

En prenant l'espérance, on obtient $r_1 = \mathbb{E}_1[r_1^{z_1}] = g(r_1)$. Ainsi, r_1 est solution de l'équation $g(s) = s$, équation qui dans $[0,1]$ admet les deux solutions s_0 et 1. Or $\mathbb{E}_1[W] = 1$ implique que $r_1 = \mathbb{P}_1(W = 0) < 1$, donc nécessairement $r_1 = s_0$ et $r_i = s_0^i$ c'est à dire que $\mathbb{P}_i(W = 0) = \mathbb{P}_i\{\zeta < \infty\}$. Mais à l'évidence, $\{\zeta < \infty\} \subset \{W = 0\}$, et donc nécessairement la variable aléatoire W est strictement positive p.s. sur $\{\zeta = \infty\}$.

□

Exemples 2.4.3. (Croissance d'une population d'insectes)

Une étude statistique a permis de montrer qu'après prise en compte de la prédation, la loi de reproduction des insectes est bien approchée par la répartition suivante :

$$p_0 = 1/6, p_1 = 1/3, p_2 = 1/3, p_3 = 1/6.$$

On vérifie aisément qu'on est bien dans le cas sur critique, avec $m = 3/2 > 1$. Autrement dit, avec une probabilité strictement positive, la taille de la population augmente exponentiellement vite avec n , et le dernier théorème permet de comprendre les fluctuations autour de la courbe $n \rightarrow (3/2)^n$.

2.4.2 Cas sous-critique

Nous nous plaçons ici dans le cas $m < 1$ et nous souhaitons étudier finement les propriétés du temps d'extinction en fonction de la condition initiale.

Proposition 2.4.1. [5] *Dans le cas sous-critique, pour tout $i, n \in \mathbb{N}^*$, on a*

$$\mathbb{P}_i(Z_n > 0) \geq \frac{i(1-m)m^{n+1}}{\sigma^2(1-m^n)+m^{n+1}(1-m)} \left(1 - \frac{(i-1)m^n}{2}\right).$$

Pour n assez grand, on a

$$\frac{m(1-m)}{\sigma^2} im^n \leq \mathbb{P}_i(Z_n > 0) \leq im^n.$$

Preuve.

Etudions tout d'abord le cas où il y a un seul ancêtre.

$$\mathbb{P}_1(Z_n > 0) \geq \frac{(1-m)m^{n+1}}{\sigma^2(1-m^n)+m^{n+1}(1-m)}. \tag{2.4.1}$$

Remarquons que $\sigma^2 = m_2 - m^2$. Nous avons

$$(\mathbb{E}_1[(Z_n \mathbb{I}_{Z_n > 0}]])^2 \leq \mathbb{E}_1[Z_n^2] \mathbb{P}_1(Z_n > 0).$$

D'où

$$\mathbb{P}_1(Z_n > 0) \geq \frac{\mathbb{E}_1[Z_n]^2}{\mathbb{E}_1[Z_n^2]} = \frac{m^{2n}}{\mathbb{E}_1[Z_n^2]}.$$

Or

$$\begin{aligned} \mathbb{E}_1[Z_n^2 | \mathcal{F}_{n-1}] &= \mathbb{E}_1\left[\sum_{k=1}^{Z_{n-1}} X_k\right]^2 \mid \mathcal{F}_{n-1}] \\ &= \mathbb{E}_1\left[\left(\sum_{k=1}^{Z_{n-1}} X_k^2 + \sum_{k,l=1}^{Z_{n-1}} X_k X_l\right) \mid \mathcal{F}_{n-1}\right] \\ &= m_2 X_{n-1} + m^2 X_{n-1} (X_{n-1} - 1). \end{aligned}$$

D'où

$$\begin{aligned} \mathbb{E}_1[Z_n^2] &= (m_2 - m^2)\mathbb{E}_1[Z_{n-1}] + m^2\mathbb{E}_1[(Z_{n-1})^2] \\ &= (m_2 - m^2)m^{n-1} + m^2\mathbb{E}_1[(Z_{n-1})^2] \\ &= \sigma^2 m^{n-1} + m^2\mathbb{E}_1[(Z_{n-1})^2] \\ &= \sigma^2(m^{n-1} + m^n + \dots + m^{2(n-1)}) + m^{2n}\mathbb{E}_1[(Z_0)^2] \\ &= \sigma^2 m^{n-1} \frac{1 - m^n}{1 - m} + m^{2n}. \end{aligned}$$

Prouvons maintenant la proposition 2.4.1. Nous avons déjà vu précédemment que

$$\mathbb{P}_i(Z_n > 0) \leq i\mathbb{P}_1(Z_n > 0) \leq i; \mathbb{E}[Z_n] = im^n.$$

Par ailleurs, nous avons

$$\begin{aligned} \mathbb{P}_i(Z_n > 0) &= 1 - \mathbb{P}_i(Z_n = 0) = 1 - (\mathbb{P}_1(Z_n = 0))^i \\ &= 1 - (1 - \mathbb{P}_1(Z_n > 0))^i. \end{aligned}$$

Nous utilisons alors le fait que $\forall k \in \mathbb{N}, z \in [0,1]$,

$$1 - (1 - z)^k \geq kz - \frac{k(k-1)}{2}z^2$$

du fait que

$$kz - (1 - (1 - z)^k) = k(k-1) \int_0^z du \int_0^u (1 - y)^{k-2} dy.$$

Nous avons alors

$$\mathbb{P}_i(Z_n > 0) \leq i\mathbb{P}_1(Z_n > 0) \left(1 - \frac{i-1}{2}\mathbb{P}_1(Z_n > 0)\right).$$

Quand n est grand, le terme prépondérant dans le terme de droite de l'inégalité (2.4.1) vaut $\frac{1-m}{\sigma^2}m^{n+1}$. □

Exemples 2.4.4. (*Extinction des baleines*)

On souhaite modéliser l'évolution d'une population de baleines (femelles) que l'on observe régulièrement dans une région du globe. L'unité de temps est un an. En 1994, des études statistiques estimaient leur nombre à 150. On suppose que l'on connaît (i.e. que l'on a pu estimer statistiquement) la probabilité qu'une baleine meure, désignons cette probabilité par $\beta \in]0,1[$. De même, supposons que l'on ait déterminé la probabilité annuelle $\alpha \in]0,1[$ qu'une baleine donne naissance à un petit.

D'après les données de la section précédente, la loi de reproduction est donnée par

$$p_0 = \beta, \quad p_1 = (1 - \beta)(1 - \alpha), \quad p_2 = \alpha(1 - \beta).$$

En 1994, des estimations ont montré que $p_0 \approx 0.6$ et $\alpha \approx 0.038$. alors

$$m = p_1 + 2p_2 = (1 - \beta)(1 + \alpha) \approx 0.976 < 1.$$

Nous sommes bien dans le cas sous-critique, avec $\sigma^2 = p_1 + 4p_2 \approx 0.095$.

Cherchons pendant combien d'années nous pouvons être certains qu'avec 99% de chance, il n'y aura pas extinction. On fait le calcul en appliquant la proposition 2.4.2. On obtient ainsi

$$\mathbb{P}(Z_n > 0 / Z_0 = 150) \geq 0.99 \implies 0.99 \leq \frac{m(1-m)}{\sigma^2} i m^n$$

c'est-à-dire

$$n \leq \frac{\log\left(\frac{0.99\sigma^2}{im(1-m)}\right)}{\log m} \leq 150.$$

Ainsi, avec 99% de chance et si il n'y a pas de changement environnemental, les baleines survivront jusqu'en 2144. Cherchons maintenant en quelle année les baleines seront éteintes, avec 99% de chance. $\mathbb{P}(Z_n = 0/Z_0 = 150) \geq 0.99$ dès que $im^n \leq 0.01$, c'est-à-dire

$n \geq 395$. Ainsi, avec 99% de chance, nous pouvons assurer que si il n'y a pas de changement environnemental, les baleines auront disparu en 2389.

La proposition ci-après permet de donner un équivalent du temps d'extinction de la population.

Proposition 2.4.2. [5] *Considérons un processus de Galton-Watson $(Z_n)_n$ sous critique ($m < 1$) et tel que $Z_0 = i$. Supposons de plus que $\sum_k k^2 q_k < \infty$ et posons $c_1 = \frac{(1-m)m}{\sigma^2}$. Soit ζ le temps d'atteinte de 0 du processus $(Z_n)_n$. Alors,*

$$\left(\frac{\ln i - \ln \ln i}{|\ln m|}\right)\left(1 - \frac{1}{ic_1}\right) \leq \mathbb{E}_i[\zeta] \leq \frac{\ln i}{|\ln m|} + \frac{1}{1-m}.$$

Corollaire 2.4.1. *Dans les mêmes conditions que ci-dessus, nous avons*

$$\mathbb{E}_i(\zeta) \sim_{i \rightarrow \infty} \frac{\ln i}{|\ln m|}.$$

Remarque 2.4.2. *Ce corollaire a des conséquences écologiques importantes: même si au temps 0, la population est extrêmement grande, elle va s'éteindre rapidement si son nombre moyen d'enfants par individu est inférieur à 1. En effet, $\ln i$ est négligeable devant i quand $i \rightarrow \infty$.*

Preuve.

Introduisons $\varphi(i) = \frac{\ln i}{|\ln m|}$ et $\psi(i) = \frac{\ln \ln i}{|\ln m|}$

$$\begin{aligned} \mathbb{E}_i(\zeta) &= \sum_n \mathbb{P}_i(\zeta > n) \\ &= \sum_{n=0}^{\varphi(i)-1} \mathbb{P}_i(\zeta > n) + \sum_{n=\lceil \varphi(i) \rceil}^{\infty} \mathbb{P}_i(\zeta > n) \\ &\leq \varphi(i) + i \sum_{n=\lceil \varphi(i) \rceil}^{\infty} m^n \\ &\leq \varphi(i) + i \frac{m^{\lceil \varphi(i) \rceil}}{1-m}, \text{ car } m < 1 \\ &\leq \frac{\ln i}{|\ln m|} + \frac{1}{1-m} \end{aligned}$$

car

$$im^{[\varphi(i)]} \leq ie^{-\frac{\ln i}{\ln m} \ln m} = 1.$$

Nous obtenons ainsi l'inégalité de droite de la proposition 2.4.2.

De plus,

$$\exp(-c_1 im^{\varphi(i)-\psi(i)}) = \exp(-c_1 m^{-\psi(i)}) = \exp(c_1 e^{-\psi(i) \ln m}) = \exp(-c_1 \ln i) = \frac{1}{ic_1}.$$

Par la proposition 2.4.1. on déduit que

$$\mathbb{P}_i(\zeta \leq n) = (1 - \mathbb{P}_1(\zeta > n))^i \leq \exp(-i\mathbb{P}_1(\zeta > n)) \leq \exp(-ic_1 m^n)$$

car $(1 - x)^k \leq \exp(-kx)$.

Ainsi, si $n \leq \varphi(i) - \psi(i)$, nous avons

$$\begin{aligned} \mathbb{P}_i(\zeta > n) &\geq \mathbb{P}_i(\zeta > \varphi(i) - \psi(i)) \\ &\geq 1 - \mathbb{P}_i(\zeta \leq \varphi(i) - \psi(i)) \\ &\geq 1 - \exp(-c_1 m^{\varphi(i)-\psi(i)}) = 1 - \frac{1}{ic_1} \end{aligned}$$

d'où

$$\begin{aligned} \mathbb{E}_i(\zeta) &\geq \sum_{n=0}^{[\varphi(i)-\psi(i)]} \mathbb{P}_i(\zeta > n) \geq (1 - \frac{1}{ic_1})([\varphi(i) - \psi(i)] + 1) \\ &\geq \left(\frac{\ln i - \ln \ln i}{|\ln m|}\right) \left(1 - \frac{1}{ic_1}\right). \end{aligned}$$

□

Revenons à l'exemple des baleines. Le corollaire 2.4.1 appliqué à $m = 0.976$ et $i = 150$ donne que $\mathbb{E}_i(\zeta) \simeq 206$. Ce qui assure les résultats du théorème 2.4.1

2.4.3 Cas critique

Théorème 2.4.3. [1] Si $m = 1$ et $\sigma^2 < \infty$ alors

1. $\lim_{n \rightarrow \infty} n\mathbb{P}(Z_n > 0) = 2\sigma^{-2}$,
2. $\lim_{n \rightarrow \infty} \frac{1}{n}\mathbb{E}[Z_n/Z_n > 0] = \frac{\sigma^2}{2}$,
3. loi($n^{-1}Z_n/Z_n > 0$) tend quand $n \rightarrow \infty$ vers la loi exponentielle de paramètre $2\sigma^{-2}$.

Preuve.

La formule de Taylor en 1 à l'ordre 2 donne

$$g(s) = s + \frac{\sigma^2}{2}(1-s)^2 + (1-s)^2\varepsilon(s)$$

avec ε bornée sur $[0,1]$ et $\lim_{s \rightarrow 1} \varepsilon(s) = 0$. Il en découle que

$$\begin{aligned} \frac{1}{1-g(s)} - \frac{1}{1-s} &= \frac{g(s) - s}{(1-g(s))(1-s)} = \frac{\frac{\sigma^2(1-s)^2}{2}(1-s)^2\varepsilon(s)}{(1-g(s))(1-s)} \\ &= \frac{\frac{\sigma^2}{2} + \varepsilon(s)}{1 - \frac{(1-s)\sigma^2}{2} - (1-s)\varepsilon(s)} = \frac{\sigma^2}{2} + \zeta(s) \end{aligned}$$

avec ζ bornée sur $[0,1]$ et $\lim_{s \rightarrow 1} \zeta(s) = 0$.

Comme $(g_k)_{k \geq 1}$ converge uniformément vers 1, il vient

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left(\frac{1}{(1-g_n(s))} - \frac{1}{(1-s)} \right) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \left(\frac{1}{(1-g_k(s))} - \frac{1}{(1-g_k(s))} \right) = \frac{\sigma^2}{2}.$$

uniformément pour tout $s \in [0,1[$. Par conséquent, on a

$$n\mathbb{P}(Z_n > 0) = n(1 - g_n(0)) = \left(\frac{1}{n} \left(\frac{1}{1-g_n(0)} - 1 + \frac{1}{n} \right) \right)^{-1} \rightarrow \frac{2}{\sigma^2}$$

De la même manière, comme $Z_n 1_{\{Z_n > 0\}} = Z_n$, on a

$$\mathbb{E}[n^{-1}Z_n/Z_n > 0] = \frac{\mathbb{E}[Z_n 1_{\{Z_n > 0\}}]}{n\mathbb{P}(Z_n > 0)} = \frac{\mathbb{E}[Z_n]}{n(1-g_n(0))} = \frac{\sigma^2}{2}.$$

Egalement de la même manière, on a pour tout $t \in \mathbb{R}_+^*$, en notant $s_n = \exp(\frac{-t}{n})$,

$$\mathbb{E}[\exp(\frac{-t}{n}Z_n)/Z_n > 0] = 1 - \frac{1-g_n(s_n)}{1-g_n(0)}$$

$$\begin{aligned}
&= 1 - \frac{1}{n(1-g_n(0))} \left(\frac{1}{n} \left(\frac{1}{(1-g_n(s_n))} - \frac{1}{1-s_n} \right) + \frac{1}{n(1-s_n)} \right)^{-1} \\
&= 1 - \frac{\sigma^2}{2} \left(\frac{\sigma^2}{2} + \frac{1}{t} \right)^{-1} = \frac{1}{1 + \frac{\sigma^2 t}{2}} \\
&= \frac{2}{\sigma^2} \int_0^\infty \frac{2}{\sigma^2} \exp(-tx) \exp\left(\frac{-2x}{\sigma^2}\right) dx.
\end{aligned}$$

2.5 Processus de Galton-Watson avec immigration

Le processus de Galton-Watson avec immigration permet de fournir une modélisation explicite de l'évolution d'une population. La notion d'immigration correspond au fait que la population de référence peut s'enrichir d'apports extérieurs: on pourra citer l'exemple d'une île où débarquent sporadiquement des individus du continent.

La chaîne de vie et de mort densité-indépendante avec immigration est décrite par

$$Z_{n+1} = \sum_{k=1}^i X_{n,k} + Y_n \quad \text{si } Z_n = i, \text{ pour } i \in \mathbb{N}. \quad (2.5.1)$$

Les variables aléatoires $(Y_n)_n$ ont la même loi η qu'une variable aléatoire Y .

La différence essentielle avec la chaîne simple (sans immigration) est que 0 n'est plus un état absorbant, puisque $p_{0ij} = \eta_j$ est strictement positif pour au moins un $j \geq 1$. Certains cas sont inintéressants :

- Si $\eta_0 = 1$, on se ramène au cas précédent.
- Si $p_0 = 1$, $Z_{n+1} = Y_n$ et la chaîne se réduit à une suite de variables aléatoires indépendantes et de même loi.
- Si $p_0 = 1$, on a $Z_{n+1} \geq Z_n + Y_n$ et la population ne peut avoir un temps d'extinction fini.

Dans la suite, nous supposons donc que

$$0 < p_0 < 1, \quad p_1 < 1, \quad \eta_0 < 1. \quad (2.5.2)$$

Nous notons par m l'espérance de la loi de reproduction. Nous supposons que chaque immigrant se comporte dès son apparition comme les autres individus et engendre donc un arbre de Galton-Watson composé de ses descendants, avec la même loi de reproduction. Les théorèmes suivants décrivent alors le comportement du processus de Galton-Watson avec immigration. Ces théorèmes généralisent les résultats du cas sans immigration, et sont plus subtils à montrer. Nous ne donnerons que des démonstrations partielles.

Théorème 2.5.1. [5] *Supposons $m < 1$. Notons $\ln^+ Y = \sup(\ln Y, 0)$. On a alors la dichotomie suivante.*

$$\begin{aligned} \mathbb{E}[\ln^+ Y] < \infty &\implies (Z_n) \text{ converge en loi.} \\ \mathbb{E}[\ln^+ Y] = \infty &\implies (Z_n) \text{ converge en probabilité vers } +\infty. \end{aligned}$$

Remarquons que ce théorème entraîne en particulier la convergence en loi de $(Z_n)_n$ si $m < 1$ et $\mathbb{E}[Y] < \infty$.

Théorème 2.5.2. [5] *Supposons $m > 1$. On a alors la dichotomie suivante.*

$$\mathbb{E}[\ln^+ Y] < \infty \implies \lim_n m^{-n} Z_n \text{ existe et est finie p.s.}$$

$\mathbb{E}[\ln^+ Y] = \infty \implies \limsup_n c^{-n} X_n = \infty$ converge en probabilité vers $+\infty$ pour toute constante $c > 0$.

Pour prouver ces théorèmes, nous allons utiliser le lemme suivant.

Lemme 2.5.1. *Soient ζ_1, ζ_2, \dots des variables aléatoires indépendantes et de même loi distribuée comme ζ . Alors pour tout $c > 1$,*

$$\begin{aligned} \mathbb{E}[\ln^+ \zeta] < \infty &\implies \sum_{n \geq 1} c^{-n} \zeta_n < \infty \text{ p.s.} \\ \mathbb{E}[\ln^+ \zeta] = \infty &\implies \limsup_n c^{-n} \zeta_n = \infty \text{ p.s.} \end{aligned}$$

Preuve. (du lemme)

Par le théorème de Borel-Cantelli, on peut vérifier que pour toute suite de v.a. V_1, V_2, \dots positives indépendantes et équidistribuées (de même loi que V),

$$\limsup_n \frac{V_n}{n} = 0 \text{ ou } \infty \text{ p.s}$$

suivant que $\mathbb{E}(V)$ est fini ou non. En effet, soit $\varepsilon > 0$. On a alors

$$\begin{aligned} \sum_n \mathbb{P}\left(\frac{V_n}{n} > \varepsilon\right) &= \sum_n \sum_{i > n\varepsilon} \mathbb{P}(V_n = i) = \sum_i \sum_{n < \frac{i}{\varepsilon}} \mathbb{P}(V_n = i) = \sum_i \sum_{n > \frac{i}{\varepsilon}} \mathbb{P}(V_n = i) \\ &= \frac{i}{\varepsilon} \mathbb{P}(V = i) = \frac{1}{\varepsilon} \mathbb{E}(V). \end{aligned}$$

Ainsi, si $\mathbb{E}[V] < \infty$ par le Théorème de Borel-Cantelli, on a $\mathbb{P}(\limsup_n \{\frac{V_n}{n} > \varepsilon\}) = 0$. Cela entraîne que la suite $(\frac{V_n}{n})_n$ converge p.s. vers 0.

Si au contraire $\mathbb{E}[V] = \infty$, et puisque les V_i sont indépendantes, alors la probabilité ci-dessus vaut 1, et donc $\frac{V_n}{n}$ tend vers l'infini p.s.

Considérons maintenant une suite $(\zeta_n)_n$ satisfaisant les hypothèses du lemme. Remarquons que si les résultats asymptotiques énoncés dans le lemme sont vrais pour la sous-suite des $\zeta_n > 0$, ils seront encore vrais pour tous les ζ_n . On pose alors $V_n = \ln^+(\zeta_n)$ et on définit $a := \ln c > 0$. Alors on a $c^{-n}\zeta_n = \exp(-n(a - \frac{V_n}{n}))$ tant que $\zeta > 1$.

Preuve. (du Théorème 2.5.1)

Remarquons que le processus de Galton-Watson avec immigration Z issu de 0 et évalué à la génération n est tel que

$$Z_n \text{ a la même loi que } \sum_{k=0}^n \xi_{n,k}$$

où les $\xi_{n,k}$ sont indépendants et pour chaque k , $\xi_{n,k}$ décrit la contribution à la génération n des immigrants arrivés à la génération $n - k$. Ainsi $\xi_{n,k}$ est la valeur à la génération k , d'un processus de Galton-Watson sans immigration lié à la loi de reproduction Y , et ne dépend donc pas de n .

Nous noterons ζ_k la condition initiale de ξ_k , c'est-à-dire le nombre d'immigrants à la génération $n - k$. Remarquons que $\xi_0 = \zeta_n$, ξ_1 est égal au nombre d'enfants des ζ_{n-1} migrants de la génération $n - 1$, etc. De plus, chaque ζ_k a même loi que Y , donc

$$\mathbb{E}[\ln^+ Y] = \mathbb{E}[\ln^+ \zeta].$$

Nous cherchons donc à déterminer si

$$Z_\infty := \sum_{k=1}^{\infty} \xi_k$$

avec ξ_k tous indépendants, est fini ou infini. On peut montrer que Z_∞ est fini p.s. ou infini p.s. Soit \mathcal{G} la tribu engendrée par toutes les variables ζ_k , $k \in \mathbb{N}$. Supposons tout d'abord que $\mathbb{E}[\ln^+ Y] < \infty$. On a alors

$$\mathbb{E}[Z_\infty | \mathcal{G}] = \sum_{k=0}^{\infty} \zeta_k m^k.$$

En effet ζ_0 a pour loi ζ , l'espérance conditionnelle de ξ_1 sachant \mathcal{G} est égale à l'espérance du nombre d'individu de la première génération issue de ζ_1 individus, c'est-à-dire $\zeta_1 m$, etc. On peut alors appliquer le lemme 2.5.1 (puisque $m < 1$), et on en déduit que l'espérance conditionnelle est finie p.s., ce qui entraîne que Z_∞ est fini p.s.

On peut montrer la réciproque, à savoir que si Z_∞ est fini p.s, alors $\mathbb{E}[\ln^+(Y)] < \infty$.

Preuve.(du Théorème 2.5.2). Considérons tout d'abord le cas où $\mathbb{E}[\ln^+ Y] = \infty$. Alors grâce au lemme 2.5.2, on sait que $\limsup_n c^{-n} Y_n = \infty$, pour tout $c > 1$. Puisque $Z_n \geq Y_n$, on en déduit le résultat.

Supposons maintenant que $\mathbb{E}(\ln^+ Y) < \infty$. On appelle \mathcal{F}_n la tribu engendrée par Z_0, Z_1, \dots, Z_n ainsi que par toutes les variables aléatoires Z_k , $k \in \mathbb{N}$. On a alors

$$\begin{aligned} \mathbb{E}\left[\frac{Z_{n+1}}{m^{n+1}} \mid \mathcal{F}_n\right] &= \frac{1}{m^{n+1}} \mathbb{E}\left[\left(\sum_{i=1}^{Z_n} X_i + Y_{n+1}\right) \mid \mathcal{F}_n\right] \\ &= \frac{Z_n}{m^n} + \frac{Y_{n+1}}{m^{n+1}}. \end{aligned}$$

Nous en déduisons que $(\frac{Z_n}{m^n})_n$ est une sous-martingale, et que, par une récurrence immédiate,

$$\mathbb{E}\left[\frac{Z_n}{m^n} \mid \mathcal{F}_0\right] = Z_0 + \sum_{k=0}^n \frac{Y_k}{m^k}, \text{ pour tout } n \geq 1.$$

Le lemme 2.5.1 entraîne alors que $(\frac{Z_n}{m^n})_n$ est une sous-martingale avec des espérances bornées, et donc elle converge p.s. vers une variable aléatoire finie p.s.

Chapitre 3

Quelques applications du processus de Galton-Watson

3.1 Introduction

La biologie va d'études très microscopiques, comme la recherche de séquences sur un brin d'ADN, l'étude des échanges moléculaires dans une cellule, l'évolution de tumeurs cancéreuses, l'invasion de parasites dans une cellule à des problèmes beaucoup plus macroscopiques concernant des comportements de grands groupes d'individus et leurs interactions (extinction de populations), ou des problèmes de génétique de populations (recherche d'ancêtres communs à plusieurs individus dans une espèce). À tous les niveaux, l'aléatoire intervient, et certains modèles peuvent aussi bien servir à décrire des problèmes de biologie posés au niveau d'individus microscopiques que de populations macroscopiques.

Le but de ce qui suit est de montrer l'importance des modèles aléatoires de type processus de branchement dans la compréhension de la biologie des populations : déplacement de cellules, croissance des bactéries, développement d'une population, évolution des espèces. L'idée de base est la suivante : même si la population semble présenter un certain nombre de caractéristiques déterministes, elle est composée d'individus dont le comportement est fondamentalement aléatoire et soumis à une grande variabilité. Ainsi, chaque individu se déplace dans une direction différente, chaque bactérie a son propre mécanisme de division cellulaire, chaque répllication de l'ADN peut engendrer une mutation. Cette variabilité individuelle est une idée fondamentale de la biologie évolutive (darwinienne).

Pour pouvoir décrire et comprendre comment la population évolue au cours du temps, il faut prendre en considération le comportement de chaque individu. La démarche du probabiliste consiste à déduire du comportement de l'individu ou de la cellule ou de la bactérie, des résultats concernant toute la population.

Les processus de branchement sont apparus comme des modèles pertinents en biologie (voir [4]) et en particulier en génétique. On trouve dans ce domaine des questions très proches de celle de la disparition des noms de famille.

Dans ce chapitre on traite deux exemples d'applications, le premier s'intéresse à la réplication d'une macro molécule d'ADN et le second aborde le problème des parasites qui croissent dans les cellules.

3.2 Exemple 1

3.2.1 Présentation du modèle

Reprenons l'exemple du deuxième chapitre (voir Kimmel-Axelrod [2], Delmas-Jourdain [3]) où nous allons calculer la probabilité d'extinction d'une macro-molécule d'ADN qui consiste en une chaîne de (ϱ) nucléotides. En une unité de temps, cette chaîne est répliquée, chaque nucléotide étant copié de façon correcte avec probabilité p et ce indépendamment des autres nucléotides.

À l'issue de la réplication, la molécule est détruite avec probabilité q ou bien donne naissance à deux molécules avec probabilité complémentaire $(1 - q)$. la probabilité de disparition de la population macro-molécules correctes est égale à celle d'extinction d'un nom de famille dans le cas où $p_0 = q$ (destruction), $p_1 = (1 - q)(1 - p^\varrho)$ (non destruction mais réplication incorrecte), $p_2 = (1 - q)p^\varrho$ (non destruction et réplication correcte) et $p_k = 0$ pour $k \geq 3$.

On introduit alors les notations et hypothèses suivantes.

On note Z_n le nombre de descendants (molécules correctes) à la n -ième génération issus d'un unique ancêtre: $Z_0 = 1$. Si $Z_n > 0$, chaque individu d'indice k compris entre 1 et Z_n a $X_{n,k}$ enfants telle que

$$Z_{n+1} = \begin{cases} 0, & \text{si } Z_n = 0; \\ \sum_{k=1}^{Z_n} X_{n,k}, & \text{Sinon.} \end{cases}$$

Les variables $(X_{n,k}, n \geq 0, k \geq 1)$ sont indépendantes et identiquement distribuées suivant la loi de reproduction $(p_j, j \in \mathbb{N}^*)$ i.e. $\mathbb{P}(X_{n,k} = j) = p_j$.

Dans ce cas, le processus de Galton-Watson est la famille des variables aléatoire $(Z_n, n \in \mathbb{N})$.

On note respectivement $m = \mathbb{E}[X_{n,k}]$, $\sigma^2 = \mathbb{E}[(X_{n,k} - m)^2]$ et $g(s) = \mathbb{E}[s^{X_{n,k}}]$, $s \in [0,1]$. leur espérance, leur variance et leur fonction génératrice communes.

D'après l'énoncé précédent, la loi de reproduction est donnée par

- $p_0 = q$: probabilité pour que la molécule soit détruite après répllication.
- $p_1 = (1 - q)(1 - p^\varrho)$: c'est le produit de la probabilité $(1 - q)$ pour que la molécule et sa copie survivent par la probabilité $(1 - p^\varrho)$ pour que l'un au moins des ϱ nucléotides ait été mal copié.
- $p_2 = (1 - q)p^\varrho$: c'est le produit de la probabilité $(1 - \varrho)$ pour que la molécule et sa copie survivent par la probabilité p^ϱ pour que chacun des ϱ nucléotides ait été copié correctement.
- $p_k = 0$, pour $k \geq 3$.

3.2.2 Etude du phénomène d'extinction

On note A l'événement "La descendance s'éteint". La probabilité d'extinction θ est donnée par

$$\theta = \mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) = g(\theta).$$

Où

$$g(\theta) = \mathbb{E}[\theta^{X_{n,j}}] = \sum_{j \in \mathbb{N}^*} \theta^j p_j.$$

On a

$$g(\theta) = q + (1 - q)(1 - p^\varrho)\theta + \theta^2(1 - q)p^\varrho.$$

Pour avoir la valeur de la probabilité d'extinction, il suffit de résoudre l'équation suivante

$$\theta = q + (1 - q)(1 - p^\varrho)\theta + \theta^2(1 - q)p^\varrho.$$

La probabilité d'extinction θ est donnée (Cf. Kimmel-Axelrod [2], Delmas-Jourdain [3]) par:

$$\theta = \frac{q}{(1-q)p^e}.$$

Remarque 3.2.1.

la position de l'espérance m par rapport à 1 joue un rôle important dans l'analyse du phénomène d'extinction.

L'espérance correspondante à notre modèle est

$$\begin{aligned} m &= g'(1) = (1-q)(1-p^e) + 2(1-q)p^e \\ &= (1-q)(1+p^e). \end{aligned}$$

Pour que la probabilité d'extinction de la population de molécules correctes soit différente de 1, il faut que $m > 1$, ce qui se récrit

$$p^e > \frac{q}{1-q}.$$

Si la probabilité q de destruction de la molécule après réplication est supérieure à $1/2$, cette condition n'est jamais vérifiée.

Dans le cas $q < 1/2$, seules les molécules dont la taille en nombre de nucléotides est inférieure à $\log(q/(1-q))/\log(p)$ peuvent survivre avec probabilité strictement positive.

3.3 Exemple 2

3.3.1 Présentation du modèle

C'est une version discrète du modèle présenté par M.Kimmel [2] afin d'étudier la probabilité d'extinction. En particulier ce modèle est régi selon une loi binomiale. Lors des divisions cellulaires, le mécanisme de transmission des parasites d'une cellule à ses deux cellules filles peut se modéliser de la manière suivante. A chaque génération, chaque parasite se multiplie indépendamment avec la même loi de reproduction. Quand les cellules se divisent chaque parasite choisit indépendamment la première cellule fille avec probabilité p , la seconde avec probabilité $1-p$.

On note Z_n le nombre de descendants (cellules correctes) à la n -ième génération issus d'un unique ancêtre: $Z_0 = 1$. Si $Z_n > 0$, chaque parasite d'indice k compris entre 1 et Z_n a $X_{n,k}$ enfants telle que

$$Z_{n+1} = \begin{cases} 0, & \text{si } Z_n = 0, \\ \sum_{k=1}^{Z_n} X_{n,k}, & \text{Sinon.} \end{cases}$$

Les variables $(X_{n,k}, n \geq 0, k \geq 1)$ sont indépendantes et identiquement distribuées suivant la loi de reproduction $(p_j, j \in \mathbb{N}^*)$ i.e. $\mathbb{P}(X_{n,k} = j) = p_j$.

Nous allons calculer la valeur de la probabilité d'extinction lorsque les $X_{n,k}$ suivent une loi binomiale de paramètres a et p , pour la comparer ensuite à la valeur numérique calculée par la suite dans le chapitre "simulations numériques".

On rappelle les expressions suivantes pour l'espérance et la variance d'une variable aléatoire X suivant une loi binomiale de paramètres a et p :

$$\mathbb{E}(X) = ap, \text{Var}(X) = apq \text{ avec } p + q = 1.$$

Selon la propriété 2.4.2, pour déterminer la probabilité d'extinction, il faut résoudre l'équation suivante:

$$g(s) = s, s \in [0,1]$$

où g est la fonction génératrice des $X_{n,k}$.

Dans le cas d'une loi binomiale de paramètres a et p , on a la fonction génératrice suivante:

$$\begin{aligned} g(s) &= \sum_{k=0}^a C_n^k p^k q^{a-k} s^k \\ &= (q + ps)^a \text{ selon le binôme de Newton.} \end{aligned}$$

Pour faciliter les calculs on se restreint au cas où $a = 2$ ensuite on résout l'équation suivante:

$$(q + ps)^2 = s$$

On développe, on obtient :

$$\begin{aligned} (q + ps)^2 - s &= 0 \\ q^2 + 2pqs + p^2s^2 - s &= 0 \\ p^2s^2 + s(2pq - 1) + q^2 &= 0 \quad (3.1) \end{aligned}$$

C'est une équation du second degré en s avec le discriminant suivant:

$$\begin{aligned}
\Delta &= (2pq - 1)^2 - 4p^2q^2 \\
&= 4p^2q^2 - 4pq + 1 - 4p^2q^2 \\
&= 1 - 4pq \\
&= 1 - 4p(1 - p) \\
&= (2p - 1)^2 \geq 0.
\end{aligned}$$

Le discriminant est nul si $p = 1/2$, par conséquent il faut distinguer deux cas :

1) $p = 1/2$: on obtient alors $\Delta = 0$ et l'équation (3.1) n'a qu'une seule racine

$$s_{p=1/2} = \frac{2pq - 1}{p^2} = 1.$$

2) $p \neq 1/2$: le discriminant est strictement positif et l'équation (3.1) admet deux racines :

$$\begin{aligned}
s_1 &= \frac{1 - 2pq + 2p - 1}{2p^2} \\
&= \frac{1 - q}{p} \\
&= 1. \quad (3.2)
\end{aligned}$$

$$\begin{aligned}
s_2 &= \frac{1 - 2pq - 2p + 1}{2p^2} \\
&= \frac{1 - (1 - p)p - p}{p^2} \\
&= \frac{1 - 2p + p^2}{p^2} \\
&= \frac{(p - 1)^2}{p^2}. \quad (3.3)
\end{aligned}$$

Lorsque $p = 1/2$, l'espérance des $X_{n,k}$ vaut 1 et la variance $1/2$, nous sommes donc dans le cas critique du théorème (2.4.1) et celui-ci affirme alors que la probabilité d'extinction vaut 1, ce qui correspond bien à la valeur calculée pour $s_{p=1/2}$.

Si $p \neq 1/2$, l'équation (3.1) admet deux racines s_1 et s_2 . Pour déterminer laquelle est la probabilité d'extinction cherchée il faut distinguer deux cas suivant la valeur de l'espérance :

1. $p < 1/2$: l'espérance est alors strictement inférieure à 1. Selon le théorème (1.3), la probabilité d'extinction vaut 1 (cas sous-critique) et c'est aussi la plus petite racine de l'équation (3.1). Donc, il faut vérifier que $s_1 < s_2$. Or on a :

$$s_2 = \left(\frac{p-1}{p}\right)^2 = \left(1 - \frac{1}{p}\right)^2$$

or

$$\begin{aligned} p < \frac{1}{2} &\Rightarrow \frac{1}{p} > 2 \Rightarrow \left|1 - \frac{1}{p}\right| = \frac{1}{p} - 1 > 2 - 1 = 1 \\ &\Rightarrow \left|1 - \frac{1}{p}\right|^2 = \left(1 - \frac{1}{p}\right)^2 > 1 \\ &\Rightarrow s_2 > s_1. \end{aligned}$$

Par conséquent, s_1 est la plus petite des deux racines et puisqu'elle vaut 1, on retrouve bien le résultat du théorème (1.3).

2) $p > 1/2$: l'espérance est alors strictement supérieure à 1 et selon la théorie, la probabilité d'extinction doit être strictement inférieure à 1 (cas sur critique). Dans ce cas, nous devons donc vérifier que

$s_2 < s_1 = 1$. Or on a :

$$s_2 = \left(\frac{p-1}{p}\right)^2 = \left(1 - \frac{1}{p}\right)^2$$

or

$$\begin{aligned} 1 > p > \frac{1}{2} &\Rightarrow 1 < \frac{1}{p} < 2 \Rightarrow \left|1 - \frac{1}{p}\right| = \frac{1}{p} - 1 < 2 - 1 < 1 \\ &\Rightarrow \left(1 - \frac{1}{p}\right)^2 < 1 \\ &\Rightarrow s_2 < s_1. \end{aligned}$$

Par conséquent, s_2 est la plus petite des deux racines et la valeur de la probabilité d'extinction du processus p vaut

$$\frac{(p-1)^2}{p^2} = s_2.$$

Chapitre 4

Simulations numériques

4.1 Résultats et analyse

En passant à la simulation, notre but est de comparer la probabilité d'extinction obtenue en pratique à la valeur obtenue théoriquement. Grâce au logiciel statistique R, nous allons générer des populations selon des processus de branchement en contrôlant les paramètres déterminants comme la loi de probabilité, l'espérance des variables aléatoires ou le nombre de générations. Nous avons choisi comme exemple directeur de générer des populations selon une loi binomiale de paramètres $a = 2$ et p .

Nous nous proposons d'exécuter notre programme de base avec plusieurs valeurs du paramètre p pour ensuite discuter des résultats obtenus. Sans modification annoncée, le programme génère 500 populations de 20 générations chacune.

4.1.1 Cas sous-critique et critique

On considère que l'espérance doit être inférieure à 1 et donc, le paramètre p doit être compris entre 0 et $1/2$.

a- Cas sous-critique :

1. On choisit d'abord $p = 0.23$; l'espérance vaut donc $2p = 0.46 < 1$. En exécutant le programme, on obtient, par exemple, la série de populations suivante:

Population	Nombre d'individus par génération
1	1 1 2 2 1 1 0
2	1 0
3	1 0
4	1 1 0
5	1 2 0
6	1 0
7	1 0

Dans cet échantillon, toutes les populations s'éteignent après un très petit nombre de générations (entre une et six générations). On exécute ensuite la fonction "compter" qui calcule le rapport du nombre de populations s'éteignant sur le nombre total de populations générées par le programme ci dessous, i.e la probabilité expérimentale d'extinction:

```
x=compter()
```

```
x
```

```
[1] 1
```

qui renvoie la valeur 1. Ce résultat correspond parfaitement à la théorie.

2. On attribue maintenant à p la valeur 0.47. L'espérance vaut alors 0.94 et est donc beaucoup plus proche de 1. Considérons à nouveau une série de sept populations engendrées selon la nouvelle valeur de p :

Population	Nombre d'individus par génération
1	1 1 2 2 1 1 1 1 1 0
2	1 2 2 2 3 2 2 1 0
3	1 1 0
4	1 0
5	1 0
6	1 2 2 1 1 1 1 0
7	1 1 1 2 1 1 1 1 2 2 1 0

Comme dans le cas (1), toutes les populations de l'échantillon s'éteignent, mais cette fois plusieurs survivent durant une dizaine de générations avant de disparaître. L'espérance étant toujours strictement inférieure à 1, la probabilité d'extinction devrait toujours valoir 1. Or, en exécutant huit fois la fonction "compter()", on obtient :

Population	Valeurs de "compter"
1	0.922
2	0.918
3	0.92
4	0.92
5	0.922
6	0.926
7	0.932
8	0.926

Si l'on prend la moyenne de ces huit valeurs, on obtient une probabilité d'environ 0.92325. Si on effectue la différence entre la valeur théorique et la valeur numérique obtenue, l'erreur est de 0.07675.

3. En choisissant $p = 0.49$, pour obtenir une espérance encore plus proche de 1 mais toujours strictement inférieure, on peut obtenir la série de populations suivante :

Population	Nombre d'individus par génération
1	1 1 2 2 3 1 2 2 2 0
2	1 1 1 1 0
3	1 2 3 2 1 0
4	1 0
5	1 1 0
6	1 1 2 1 1 0
7	1 2 2 3 3 3 3 4 2 3 2 1 1 2 3 2 2 2 1 1

La dernière population survit même au-delà des 20 générations considérées. Cependant, il ne reste qu'un seul individu à la vingtième génération et on peut donc supposer que la population va s'éteindre après quelques générations. Dans ce cas, le programme retourne une probabilité moyenne de 0.8725, ce qui correspond à une erreur moyenne de 0.1275.

Cette probabilité correspond au quotient des populations qui s'éteignent sur le nombre total de populations généré. Par conséquent, cela signifie que des populations survivent au-delà des 20 générations considérées et ne s'éteignent qu'après, ce qui fausse le calcul de la probabilité d'extinction. On peut supposer qu'on peut diminuer l'erreur et donc améliorer la probabilité en augmentant le nombre de générations considérées. Par exemple, en augmentant de 30, le nombre de générations considérées, on obtient une probabilité moyenne de 0.966, soit une erreur de 0.034, ce qui est déjà plus acceptable.

Cependant, on trouve encore des populations qui survivent à 50 générations, comme l'exemple suivant

```
[1] 1 2 4 5 5 6 6 7 7 8
[11] 11 14 17 19 20 25 27 25 27 24
[21] 22 26 32 27 31 33 25 21 21 21
[31] 23 28 26 25 24 21 22 24 25 30
[41] 26 25 24 23 24 27 31 31 28 25
```

A l'allure du graphique de cette population, on peut supposer qu'elle va s'éteindre plus ou moins lentement.

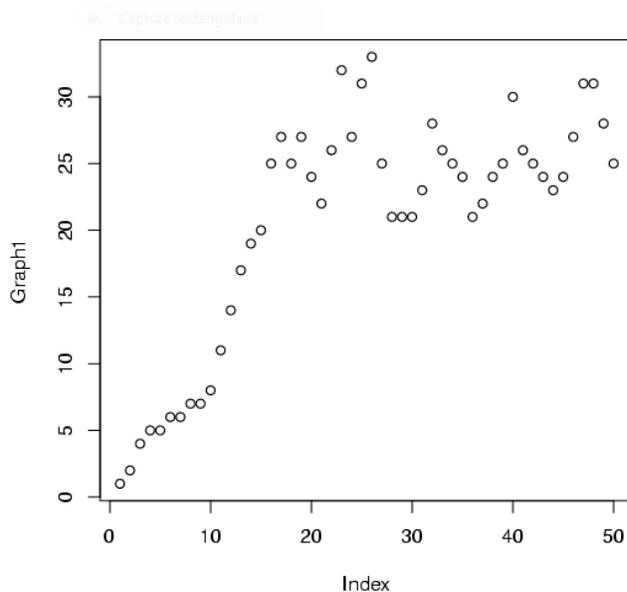


FIG. 4.1 – *Graphe 1*

On peut également déduire de ces exemples que plus l'espérance s'approche de 1 et par conséquent plus p s'approche de $1/2$, plus grand est le nombre de générations durant lesquelles la population survit, même si finalement celle-ci finit toujours par s'éteindre, et donc plus grand est le nombre de générations à considérer pour minimiser l'erreur.

b- Cas critique

Dans ce cas, $p = 1/2$ et l'espérance est égale à 1 et la variance est strictement positive. La théorie affirme alors que la probabilité d'extinction vaut 1.

En considérant le même programme que dans le cas sous-critique mais avec $p = 1/2$, on se rend compte que le problème de l'erreur subsiste voir s'aggrave. En effet, si nous considérons 20 générations, on trouve une probabilité moyenne de 0.848. Comme avant, cette erreur provient du nombre insuffisant de générations considérées. En passant à 50 générations, la probabilité moyenne devient 0.932 ce qui correspond à une erreur de 0.068. Pour avoir une erreur d'environ 0.01, il faut encore augmenter le nombre de générations. En essayant de trouver le nombre de générations qui minimise l'erreur, on s'aperçoit que celle-ci diminue moins vite dans ce cas que dans le cas précédent : en effet, en passant à 100 générations, la probabilité moyenne vaut 0.954 et donc les générations ont doublé mais l'erreur n'a pas été divisée par 2 (on passe de 0.068 à 0.046).

En résumé, on constate que plus p s'approche de la valeur critique $1/2$, plus l'erreur est grande si l'on ne considère que 20 générations. Autrement dit, le nombre de populations survivant au-delà des 20 générations va en croissant avec p . Pour minimiser cette erreur, il faut donc augmenter le nombre de générations prises en compte par le programme, pour intégrer au calcul ces populations s'éteignant après la limite de 20 générations. Comme le nombre de ces populations va en croissant avec p , plus on s'approche de $1/2$, plus il faut augmenter le nombre de générations considérées, comme le montre le cas critique.

Remarque 4.1

Remarquons que le fait d'augmenter le nombre de générations n'influence que peu voire pas du tout sur le temps d'exécution du programme, puisque les populations étant vouées à disparaître, le nombre d'individus à chaque génération n'explose pas et le nombre d'opérations est ainsi limité.

4.1.2 Cas sur-critique

On considère à présent le paramètre p compris strictement entre $1/2$ et 1, ainsi l'espérance est strictement supérieure à 1.

1. $p = 0.55$: en exécutant le programme plusieurs fois, on obtient par exemple, la série de populations suivante :

Population	Nombre d'individus par génération
1	1 2 2 3 3 1 1 2 3 3 5 5 3 4 4 4 2 2 2 2
2	1 1 2 1 0
3	1 2 2 2 2 2 3 4 3 4 3 4 7 10 8 8 8 8 9 9
4	1 1 2 3 3 1 0
5	1 1 1 1 0
6	1 1 1 1 1 1 0
7	1 1 1 2 3 4 5 8 9 10 11 14 16 17 18 22 22 27 34 45
8	1 1 0
9	1 1 1 1 1 1 0

Parmi les populations qui survivent, on peut comparer les graphiques de 2 et 3. En effet, elles dépassent toutes les deux les 20 générations, mais dans le cas 2, on peut supposer qu'elle va s'éteindre rapidement, alors qu'à la vue du graphe exponentiel de 3 on peut supposer qu'elle va survivre indéfiniment.

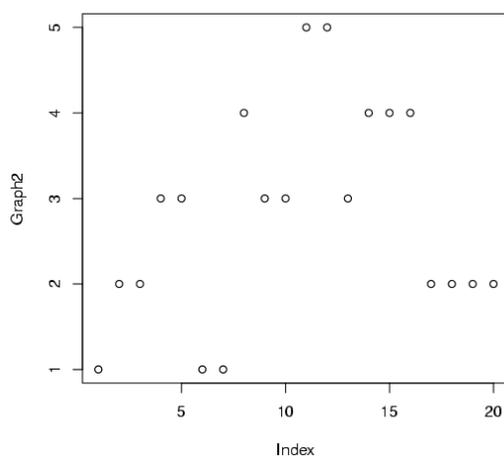


FIG. 4.2 – Graphe 2

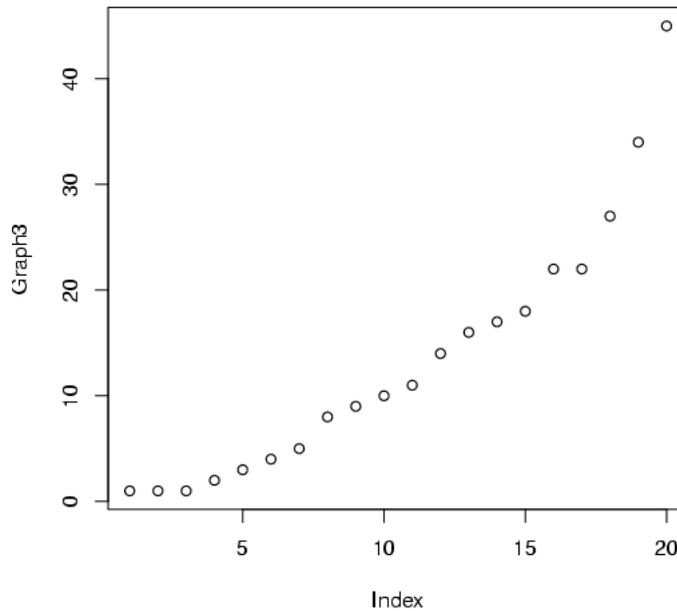


FIG. 4.3 – Graphe 3

Si l'on exécute la fonction "compter", on obtient une probabilité d'extinction numérique moyenne (sur huit exécutions) de 0.64325. Théoriquement, en résolvant l'équation (3.3) pour $p = 0.55$, on devrait trouver :

$$\theta = \frac{(p - 1)^2}{p^2} \cong 0.66942.$$

L'erreur vaut donc 0.02617. En considérant 30 générations, on obtient une probabilité de 0.65366 et une erreur moyenne de 0.01576.

2. $p = 0.9$, en exécutant le programme plusieurs fois, on obtient par exemple, la série de populations suivante :

Population	Nombre d'individus par génération
1	1 2 3 5 8 13 24 41 75 130 232 422 769 1380 2500 4489 8130 14645 26321 47320
2	1 1 2 4 8 16 29 55 99 179 319 576 1042 1882 3391 6079 10914 19607 35420 63786
3	1 2 3 6 11 21 40 74 135 250 459 835 1504 2692 4877 8819 15837 28533 51375 92497
4	1 1 1 2 4 5 9 14 23 46 84 155 279 500 887 1583 2858 5172 9296 16774
5	1 2 3 4 6 11 18 35 62 117 208 375 659 1199 2139 3806 6853 12376 22361 40303
6	1 2 2 3 6 11 20 37 61 116 209 375 678 1223 2189 3935 7085 12724 22930 41273
7	1 2 3 6 10 18 28 51 94 168 302 544 976 1752 3155 5640 10118 18192 32718 58936

Dans cet échantillon, toutes les populations survivent. De plus, le nombre d'individus croît exponentiellement, comme le montre le graphe de la dernière population de l'échantillon.

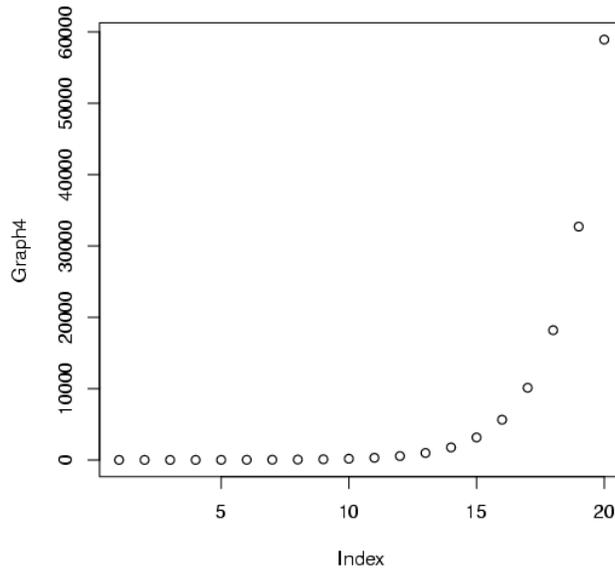


FIG. 4.4 – *Graphe 4*

Dans ce cas, le probabilité d'extinction calculée par le programme vaut 0.012. La valeur théorique est:

$$\theta = \frac{(p-1)^2}{p^2} \cong 0.01235.$$

L'erreur est donc minime même sur 20 générations.

Conclusion générale

Le travail de ce mémoire a porté sur l'étude des processus de Galton-Watson et applications .

En premier lieu, nous avons introduit les grandes classes des processus de branchement, markoviens à temps discret dont le processus de Galton-Watson avec immigration et sans immigration, ainsi que les processus de branchement markoviens à temps continu, tels que le processus de Lévy, le processus de Crump-Mode-Jagers et le processus de Bellman-Harris. Par la suite, nous avons abordé la modélisation des évolutions de population et le calcul des probabilités d'extinction à l'aide du processus de Galton-Watson et pour mieux comprendre tout cela, on a donné deux applications où nous nous sommes intéressés à la disparition de macro-molécules d'ADN, pour le premier exemple et à la probabilité d'extinction de la population de parasite pour le second. A la fin, on a vérifié le bien-fondé de la théorie à travers des simulations numériques sur le second exemple étudié au chapitre 3 en utilisant le logiciel R.

Comme perspectives de ce travail, on peut essayer de proposer d'autres modèles aléatoires prenant en compte d'autres aspects dans la dynamique de la population (immigration, survie pendant plusieurs générations,...).

Des applications numériques sur des données réelles peuvent être également envisagées.

Annexe A

Code source

Voici l'un des codes sources utilisés pour nos simulations, celui-ci concerne une modélisation de générations où les $X_{n,k}$ suivent une loi de probabilité binomiale de paramètres $\mathbf{a} = \mathbf{2}$ et \mathbf{p} .

```
p <- 0.8
n <- 2
m <- 20
k <- 500
```

On fixe la valeur de \mathbf{p} à $\mathbf{0.8}$, \mathbf{m} est le nombre de générations considérées, on prend $\mathbf{m} = \mathbf{20}$ et \mathbf{k} le nombre de populations générées, on choisit $\mathbf{k} = \mathbf{500}$.

```
gen <- fonction(z)
{
x <- rbinom(z,n,p)
gen <- sum(x)
}
gen
```

“**gen**” est la fonction qui nous permet de fabriquer une génération, elle ne prend qu’un seul argument (\mathbf{z}). Elle génère \mathbf{z} valeurs suivant une loi binomiale de paramètres \mathbf{a} et \mathbf{p} grâce à la fonction **rbinom**(\mathbf{z} , \mathbf{a} , \mathbf{p}) et les stocke dans les composantes de \mathbf{x} , puis retourne la somme des composantes de \mathbf{x} , ce qui représente le nombre total d’individus d’une génération.

```

genpop <- function()
{
Z <- 1
for( i in 2:m)
{
Z <- c(Z,gen(Z[length(Z) ]))
if(Z[length(Z) ]==0) break]
}
if(Z[length(Z) ]!=0) Z <- c(Z,1)]
genpop <- Z
}

```

“**genpop**” est la fonction qui nous permet de fabriquer une population constituée de **m** générations. On fixe **Z** = **1**, puis on a une boucle de 2 à **m**, à chaque itération, la fonction ajoute une composante au vecteur **Z**, cette composante est donnée par la fonction “**gen**” qui prend pour argument la dernière composante de **Z** c’est à dire le nombre d’individus de la dernière génération. Les composantes de **Z** sont donc le nombre d’individus(parasites) de chaque génération. Ensuite la fonction effectue un test, si la dernière génération ne compte aucun parasite, la boucle s’arrête; sinon, après la boucle on ajoute la composante 1 au vecteur **Z** ce qui nous sera utile pour la fonction suivante. Cette fonction retourne **Z**.

```

compter <- function()
{
surv <- 0
for(i in 1:k)
{
pop <- genpop()
surv <- surv + pop[length(pop) ]
}
prob <- (k-surv)/k
compter <- prob
}

```

“**compter**” est la fonction qui nous retourne la probabilité expérimentale (différente de sa valeur théorique) d’extinction de la population générée suivant la loi de probabilité donnée, cette fonction ne prend pas d’argument. On fixe **surv** à 0. La fonction exécute une boucle de 1 à **k**. A chaque itération on fabrique une population (composée de **m** générations), et on ajoute la dernière composante de “**genpop**”, c’est à dire 0 si la population s’est éteinte et 1 si elle a survécu. **surv** est donc le nombre de populations ayant survécu, et **k-surv** le nombre de celles qui se sont éteintes. Enfin la fonction retourne la probabilité d’extinction d’une population.

Remarque 1 Si on souhaite modifier la loi de distribution, il suffit de modifier le corps de la fonction “**gen**” en remplaçant la fonction **rbinom(z, n, p)** par une autre loi de distribution, par exemple la loi de Poisson de paramètre **Lambda** :

```
gen <- function(z)
{
x <- pois(z,Lambda)
gen <- sum(x)
}
```

Bibliographie

- [1] **Jacques Istas**, Mathematical Modeling for the Life Sciences, Département IMSS BSHM Université Pierre Mendès-France, Vol.34, Springer-Verlag 2000.
- [2] **M. Kimmel, D.E. Axelrod**, Branching Processes in Biology, Springer 2002.
- [3] **J.F. Delmas, B. Jourdain**, Modèles aléatoires: applications aux sciences de l'ingénieur et du vivant, Springer 2006.
- [4] **K.B. Athreya, P.E. Ney**, Branching Processes, Springer 1972.
- [5] **Sylvie Méléard**, Modèles aléatoires en Ecologie et Evolution.
- [6] **Charles Robert Darwin**, On the Origin of Species by Means of Natural Selection. John Murray, London 1859.
- [7] **Jyy-I Hong**, Coalescence in Bellman-Harris and multi-type branching processes, Iowa State University 2011.
- [8] **Peter Jagers**, Branching processes with biological applications, London, 1975.
- [9] **Olle Nerman**, On the convergence of supercritical general branching processes. Z. Wahrsch. Verw. Gebiete, 1981.
- [10] **Peter Jagers and Olle Nerman**, Limit theorems for sums determined by branching and other exponentially growing processes. Stochastic Process.
- [11] **Amaury Lambert**, The contour of splitting trees is a Lévy process, 2008.
- [12] **Mathieu RICHARD**, Arbres, Processus de branchement non markoviens et Processus de Lévy, Thèse de doctorat, École Doctorale Paris Centre, Université Pierre et Marie Curie, 2011.
- [13] **N. Champagnat, S. Méléard**, Polymorphic evolution sequence and evolutionary branching, Probability Theory and Related Fields 151, 45-94 (2011).
- [14] **F. Comets, T. Meyre**, Calcul stochastique et modèles de diffusion, Dunod, 2006.
- [15] **U. Dieckmann, M. Doebeli**, On the origin of species by sympatric speciation. Nature, 400 :354-357, 1999.
- [16] **S.N. Ethier, T.G. Kurtz**, Markov Processes, Characterization and Convergence, New York, 1986.

- [17] **W. Feller**, An introduction to Probability Theory and its Applications, 2 Vol. Wiley, 1957.
- [18] **C. Graham**, Chaînes de Markov, Dunod, 2008.
- [19] **P. Haccou, P. Jagers, V.A. Vatutin**, Branching Processes: Variation, growth and extinction of populations, Cambridge University Press, 2005.
- [20] **N. Ikeda, S. Watanabe**, Stochastic Differential Equations and Diffusion Processes, 2nd Edition, North-Holland Publishing Company, 1989.
- [21] **J. Istas**, Introduction aux modélisations mathématiques pour les sciences du vivant. Springer, 2000.
- [22] **J. Jacod, P. Protter**, L'essentiel en théorie des probabilités. Cassini, 2002.
- [23] **I. Karatzas, S.E. Shreve**, Brownian Motion and Stochastic Calculus, Second Edition Springer, 1998.
- [24] **S. Karlin, H.M. Taylor**, A First Course in Stochastic Processes, Academic Press, 1975.
- [25] **S. Karlin, H.M. Taylor**, A Second Course in Stochastic Processes, Academic Press, 1981.
- [26] **S. Méléard**, Aléatoire: Introduction à la théorie et au calcul des probabilités. Editions de l'Ecole Polytechnique, 2010.
- [27] **J. Neveu**, Martingales discrètes, Masson, 1972.
- [28] **S. Wright**, Evolution in mendelian populations, Genetics, 1931.
- [29] **M.A. Yaglom**, Certain limit theorems of the theory of branching random processes, Reports of the Academy of Sciences of USSR, 1947.
- [30] **S. Tavaré, D.J. Balding, R.C. Griffiths and P. Donnelly**, Inferring coalescence times from DNA sequence data, Genetics, 1997.