

# Rapport de stage de M1

*Autour du comportement asymptotique du processus de Galton-Watson et de ses généralisations*

Jonathan BADIN

## Introduction

J'ai effectué mon stage de M1 dans le cadre de ma deuxième année d'étude à l'ENS à l'université du Québec à Montréal (UQÀM) sous la supervision de Dante Mata-López et d'Hélène Guerin. L'objectif du stage était de découvrir le processus de Galton-Watson et une de ses généralisations, le processus de Galton-Watson à environnement variable. Avec comme point de mire, le récent article [CTJP23] qui étudie la vitesse de convergence d'un théorème de convergence classique, le théorème de convergence de Yaglom dans le régime critique, pour un environnement variable. Avant d'en arriver là il a fallut revoir les résultats fondamentaux du processus de Galton-Watson classique afin de partir sur des bonnes bases et de mieux comprendre les mécanismes du processus. Une étape fondamentale fut l'article [LPP95] et leur construction de l'arbre biaisé. J'ai alors commencé par étudier la vitesse de convergence du processus conditionné dans le cas d'un environnement constant en suivant la démarche de [PR11] qui repose sur la méthode de Stein. Il a fallut me mettre à jour sur les fondements de cette méthode que j'ai résumées dans ce rapport. Pour terminer, je me suis tourné vers l'environnement variable en commençant par l'article [Ker20] qui généralise les résultats fondamentaux du cas univarié et finalement sur l'article [CTJP23] qui réutilise la méthode de [PR11] pour obtenir des bornes sur la vitesse de convergence en environnement variable.

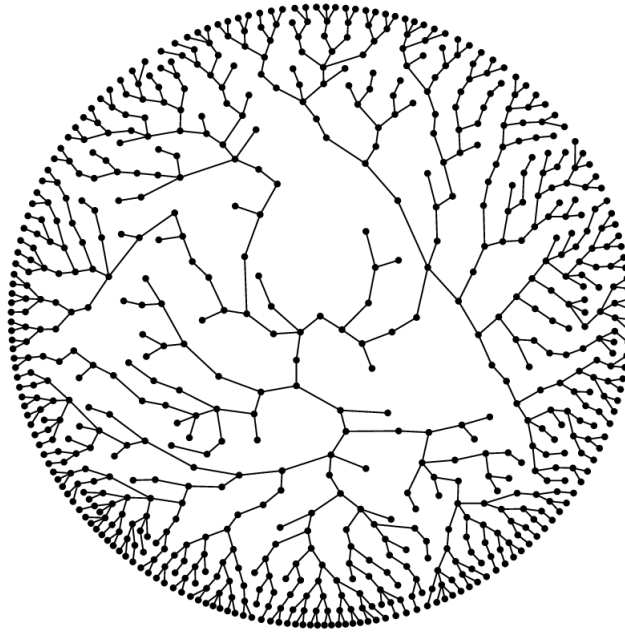


FIGURE 1 – Illustration du processus de Galton-Watson. La distribution des descendants est binomiale, de paramètres  $n = 3$  et  $p = 0.4$ . Le temps s'écoule de haut en bas. La figure est provient de Martingales et calcul stochastique, Nils Berglund, 2013, figure 1.5.

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Comportement asymptotique du processus de Galton-Watson classique</b>	<b>3</b>
1.1 Fonction génératrice . . . . .	5
1.2 Convergence du processus . . . . .	6
1.3 Comportement vers l'explosion . . . . .	7
1.4 Comportement vers l'extinction . . . . .	8
Annexe - Arbre de Galton-Watson . . . . .	9
<b>2 L'arbre biaisé</b>	<b>12</b>
2.1 Immigration . . . . .	13
2.2 Critère $X \log(X)$ . . . . .	13
2.3 Théorème de convergence de Yaglom critique . . . . .	14
Annexe - Arbre de Galton-Watson biaisé . . . . .	16
<b>3 Vitesse de convergence vers la limite de Yaglom</b>	<b>18</b>
3.1 Méthode de Stein . . . . .	18
3.2 Vitesse de convergence . . . . .	20
Annexe - Reste dans l'estimation de Kolmogorov . . . . .	21
<b>4 Processus de Galton-Watson à environnement variable</b>	<b>24</b>
4.1 Définition du processus . . . . .	24
4.2 Fonctions de formes . . . . .	25
4.3 Convergence . . . . .	26
4.4 Extinction . . . . .	27
4.5 Limite de Yaglom . . . . .	28
4.6 Vitesse de convergence . . . . .	29
<b>Conclusion</b>	<b>31</b>
<b>Annexes - Démonstrations des théorèmes</b>	<b>32</b>

# 1 Comportement asymptotique du processus de Galton-Watson classique

On commence par introduire le processus de Galton-Watson pour ensuite se tourner vers l'étude asymptotique du processus. On commence par analyser la convergence du processus pour ensuite étudier plus en détails le comportement asymptotique du processus.

## Introduction

A la fin du 19<sup>ème</sup> siècle, l'Angleterre victorienne se préoccupe de la disparition des patronymes de ses grandes familles.

*The decay of the families of men who occupied conspicuous positions in past times has been a subject of frequent remark, and has given rise to various conjectures. It is not only the families of men of genius or those of the aristocracy who tend to perish, but it is those of all with whom history deals, in any way ... The instances are very numerous in which surnames that were once common have since become scarce or have wholly disappeared.*

Ainsi écrit le polymathe F.GALTON qui se penche sur le sujet. Refusant d'accepter pour argent comptant l'idée répandue à l'époque selon laquelle cette disparition est due à une diminution de la fertilité chez les familles distinguées, il suggère d'étudier la probabilité qu'une famille ordinaire disparaisse à partir de données statistiques sur la population anglaise.

*The tendency is universal, and, in explanation of it, the conclusion has been hastily drawn that a rise in physical comfort and intellectual capacity is necessarily accompanied by diminution in "fertility". On the other hand, M.ALPHONSE DE CANDOLLE has directed attention to the fact that, by the ordinary law of chances, a large proportion of families are continually dying out, and it evidently follows that, until we know what that proportion is, we cannot estimate whether any observed diminutions of surnames among the families whose history we can trace, is or is not a sign of their diminished "fertility".*

Rencontrant des difficultés dans la résolution de ce problème, il sollicite l'aide d'autres mathématiciens et pose son problème dans le journal "*The Educational Times*"<sup>1</sup> sous la forme suivante.

*A large nation, of whom we will only concern ourselves with adult males,  $N$  in number, and who each bear separate surnames, colonise a district. Their law of population is such that, in each generation,  $a_0$  per cent of adult males have non male children who reach adult life;  $a_1$  have one such male child;  $a_2$  have two; and so on up to  $a_5$  who have five. Find (1) what proportion of the surnames will have become extinct after  $r$  generations; and (2) how many instances there will be of the same surname being held by  $m$  persons.*

Le révérent H.W.WATSON répond avec une solution<sup>2</sup>. Suite à cela, ils écrivent ensemble un papier intitulé *On the probability of the extinction of families* (1874) dont sont tirés les extraits précédents. Aujourd'hui, le sujet porte leurs deux noms. Mais bien d'autres mathématiciens ont indépendamment découvert le problème et sa solution. Dès 1845, le mathématicien français I.J BIENYAMÉ entreprit des travaux similaires. Mais également A.A COURNOT en 1847 dans le livre *de la correspondance de l'algèbre et de la géométrie* sous la formulation d'un problème de jeu. Par la suite, le sujet continua de se développer avec notamment les travaux du biologiste R.A FISHER (~ 1920) qui l'appliqua à l'étude de la survie d'un gène mutant, SEMENOFF (1935) via les réactions chimiques en chaîne, SHOCKLEY et PIERCE (1938) via les multiplicateurs d'électrons mais c'est surtout après les années 1940 avec le développement de la bombe atomique que le modèle gagna en

1. Numéro 143 datant du 1<sup>er</sup> mars 1873 problème 4001.

2. Numéro 148 datant du 1<sup>er</sup> août 1873.

intérêt. On citera notamment les travaux de S.ULAM et D.HAWKINS durant le *projet Los Alamos*. Aujourd'hui, on formule le modèle de la façon suivante :

**Définition 1.** Le processus de Galton-Watson de loi  $f$  est le processus  $(Z_n : n \geq 0)$  défini par son premier terme  $Z_0 = 1$  et par la récurrence

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n} \quad (1)$$

où les  $X_{i,n}$  sont des variables aléatoires indépendantes de loi  $f$ .

**Hypothèse H.** Afin d'exclure les cas triviaux on suppose que la loi de reproduction  $f$  est telle que

$$\begin{aligned} 0 < f[0] < f[0] + f[1] < 1, \\ f[z] &\neq \delta_z, \forall z \in \mathbf{N}. \end{aligned} \quad (H)$$

Le problème est alors d'étudier le *comportement asymptotique* du processus : atteint-il la valeur 0 en y restant indéfiniment, sinon vers quelles valeurs converge-t-il et comment ? Selon le point de vue historique, on utilise toujours le vocabulaire des populations. Nous parlerons donc d'individus, d'enfants, de générations etc... Ainsi,  $n$  est la génération,  $X_{i,n}$  est le nombre d'enfants de l'individu  $i$  à la génération  $n$  et  $Z_n$  le nombre d'individu de la population à la génération  $n$ . Cela permet de se représenter le processus comme un arbre aléatoire : l'arbre généalogique de sa population (voir figure 1 pour un exemple). Les détails quand au formalisme mathématique des arbres de Galton-Watson est laissé en annexe. Ce qu'il faut juste retenir est qu'il existe une mesure de probabilité  $\text{GW}_f$  sur l'ensemble des arbres tel que si  $Z_n$  est la fonction mesurable qui associe à chaque arbre, le nombre de sommets à la profondeur  $n$  alors sous  $\text{GW}_f$ ,  $(Z_n | n \geq 0)$  forme un processus de Galton-Watson de loi de reproduction  $f$ .

### Exemple. Division cellulaire

Une cellule se reproduit via la processus de division cellulaire (*mitose*). A partir d'une cellule mère, ce procédé permet d'obtenir deux cellules génétiquement identiques. Mais, il peut arriver, notamment via les interactions extérieures, que certaines cellules meurent. On peut représenter cette situation par un processus de Galton-Watson de loi de reproduction  $f$  prenant exclusivement deux valeurs : 0 et 2. Nous parlerons de *branchement binaire*. La figure 2 présente des simulations bien choisies du processus de Galton-Watson à branchement binaire pour divers paramètres.

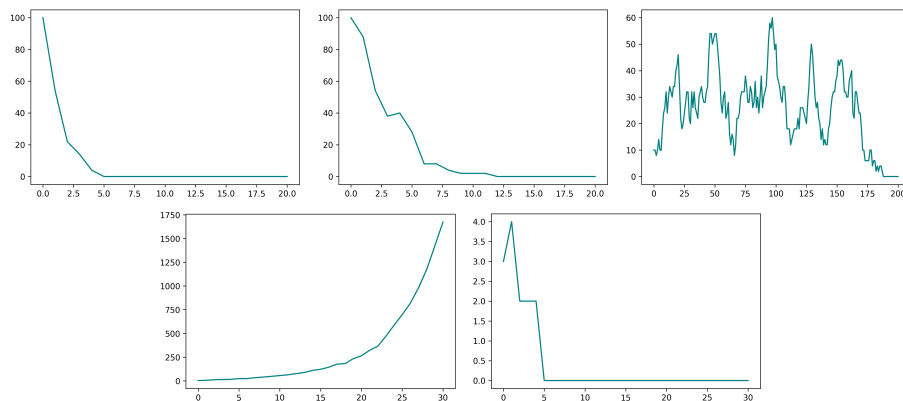


FIGURE 2 – Simulations d'un processus de Galton-Watson à branchement binaire  $f = (1-p)\delta_0 + p\delta_2$  avec respectivement  $p = 0.3, 0.4, 0.5, 0.6$  et  $0.7$  et pour une population initiale fixe respectivement 100, 100, 10, 3 et 3.

## 1.1 Fonction génératrice

La propriété fondamentale du processus de Galton-Watson est l'expression de sa fonction génératrice. Dans toute la suite, nous utiliserons le même symbole pour désigner la loi de probabilité et la fonction génératrice associée, pour éviter quelques confusions nous emploierons des parenthèses lorsqu'il s'agit de la fonction génératrice et des crochets lorsqu'il s'agit de la loi de probabilité. Ainsi,

$$f(s) = \sum_{k=0}^{\infty} f[k]s^k.$$

**Proposition 1.** *Propriété fondamentale*

La fonction génératrice de  $Z_n$  est

$$\mathbf{E}[s^{Z_n}] = f \circ \dots \circ f =: f^n. \quad (2)$$

Bien que la formule 2 établisse une expression explicite de la fonction génératrice du processus, cette expression est en pratique inaccessible, l'itération d'une fonction même simple menant à des calculs inextricables. Il y a bien un cas où l'on peut calculer explicitement  $f^n$ , et c'est bien le seul, c'est le cas où  $f(s)$  est une homographie. Dans ce cas  $f$  est de la forme suivante

$$\begin{aligned} f[0] &= a, \\ f[k] &= (1-a)b(1-b)^{k-1} \text{ pour } k \geq 1. \end{aligned}$$

On peut voir cette loi comme une composée d'une loi géométrique et d'un dirac, on parle de loi *linéaire fractionnaire*.

Dans tous les cas, la formule précédente permet de récupérer les premiers moments du processus.

**Corollaire 2.** *Si  $f$  admet un moment d'ordre  $k$ , alors  $Z_n$  également et,*

$$\mathbf{E}[Z_n(Z_n - 1) \cdots (Z_n - k + 1)] = \frac{d^k}{ds^k} f^n(s). \quad (3)$$

En particulier,

- si  $f$  a une espérance  $\mu$  alors  $Z_n$  aussi et,

$$\mathbf{E}[Z_n] = \mu^n. \quad (4)$$

- si  $f$  a de plus une variance  $\sigma^2$  alors  $Z_n$  aussi et,

$$\mathbf{Var}(Z_n) = \begin{cases} \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1} & \text{si } \mu \neq 1, \\ n\sigma^2 & \text{si } \mu = 1. \end{cases} \quad (5)$$

Nous supposerons toujours que  $f$  admet une moyenne que l'on désignera en toute circonstance par la lettre  $\mu$ . Suivant la valeur de  $\mu$ , on s'attend à voir des comportements radicalement différents. On voit bien se dessiner trois régimes :

- le régime  $\mu < 1$  dit *sous-critique*. Dans ce cas, l'espérance du processus tend géométriquement vers 0. Avec l'inégalité de Markov,

$$\mathbf{P}(Z_n > \varepsilon) \leq \frac{1}{\varepsilon} \mathbf{E}[Z_n] = \frac{1}{\varepsilon} \mu^n$$

on voit que le processus s'éteint presque sûrement *i.e.*  $Z_n \rightarrow 0$  presque sûrement<sup>3</sup> c'est à dire comme  $Z_n$  prend des valeurs entières que presque sûrement  $Z_n = 0$  à partir d'un certain rang.

---

3. Rappelons qu'il découle du lemme de Borel-Cantelli que si pour tout  $\varepsilon > 0$ ,  $\sum_{n=1}^{\infty} \mathbf{P}(|X_n - X| > \varepsilon) < \infty$  alors  $X_n$  converge presque sûrement vers  $X$ .

- le régime  $\mu = 1$  dit *critique*. Dans ce cas, l'espérance du processus est constante, mais la variance tend vers l'infini. On s'attend donc à avoir des générations de grandes tailles, mais qui finiront par s'éteindre.
- le régime  $\mu > 1$  dit *sur-critique*. Dans ce cas, l'espérance du processus tend géométriquement vers l'infini. On s'attend à avoir une population qui explose. Par contre, on ne peut espérer avoir dans le régime sur-critique, une explosion presque sûre de façon duale au cas sous-critique. Car sauf cas à part, l'extinction est toujours possible et est irréversible. Par exemple, dans le cas d'une loi géométrique i.e.  $f[k] = (1-p)p^{k-1}$  où  $p \in (0,1)$ , l'espérance de  $f$  est  $\frac{1}{p} > 1$  mais on a toujours une probabilité  $p$  de s'éteindre à la première étape. Plus la population est grande, moins le processus a de chance de s'éteindre. En quelque sorte, dans le régime sur-critique, le développement des premières générations est décisive ; si la population se développe, elle a des chances de survivre, sinon elle risque de s'éteindre.

**Exemple.** *Régimes du branchement binaire*

Reprenons le cas d'un branchement binaire  $f(s) = (1-p) + ps^2$ . On calcule  $f'(1) = 2p$  de sorte que :

- $p < 1/2$  correspond au régime sous-critique ;
- $p = 1/2$  correspond au régime critique ;
- $p > 1/2$  correspond au régime sur-critique.

**Remarque.** *Modèle Malthusien*

On peut faire le parallèle entre l'évolution de la moyenne (4) et le modèle de croissance malthusien <sup>4</sup>. Ce modèle de d'évolution de population basique décrit la dynamique d'une population selon

$$\frac{dx(t)}{dt} = \alpha x(t)$$

où  $\alpha$  désigne le taux de croissance de la population. Cette équation se résout simplement en

$$x(t) = x_0 e^{\alpha t}$$

où  $x_0$  désigne la population initiale.

## 1.2 Convergence du processus

Le processus de Galton-Watson possède la propriété de Markov ; c'est une chaîne de Markov homogène de transition

$$P(i, j) = f^{*i}(j)$$

où l'étoile désigne la convolution. L'état 0 est absorbant et sous l'hypothèse 0, tous les autres états sont transitoires, étant connectés à 0 :  $P(i, 0) = f^{*i}(0) = f^i(0) > 0$ . Comme une chaîne de Markov, ne passe qu'un nombre fini de fois en chaque état transitoire, soit le processus est absorbé par 0 on parle d'*extinction*, soit le processus diverge vers  $+\infty$ , on parle d'*explosion*. On a ainsi une dichotomie extinction-explosion, le processus ne peut se stabiliser sur une valeur fini non nulle.

On cherche maintenant à quantifier la dichotomie précédente. Pour cela on introduit la *probabilité d'extinction*

$$q = \mathbf{P}(\lim Z_n = 0)$$

Commençons par remarquer que  $\{\lim Z_n = 0\}$  est l'union croissante des évènements  $\{Z_n = 0\}$  de sorte que

$$q = \lim_{n \rightarrow \infty} \mathbf{P}(Z_n = 0) = \lim_{n \rightarrow \infty} f^n(0).$$

On en déduit que  $q$  est un point fixe de  $f$ . Pour déterminer la probabilité d'extinction, il convient donc de s'intéresser aux points fixes de  $f$ . On est donc ramener à une étude de fonction. La propriété

4. En l'honneur de Thomas Robert MALTHUS, qui écrit un l'influent livre "*An Essay on the Principle of Population*" (1798).

fondamentale en ce sens est que  $f$  est *strictement convexe* sur  $(0, 1)$ . C'est une propriété générale des fonctions génératrices d'être convexe, la stricte convexité découle quant à elle de l'hypothèse 0. Comme de plus  $f$  est une fonction génératrice, 1 est un point fixe de  $f$  et avec la propriété de convexité, on voit que  $f$  admet au plus un autre point fixe. Et l'existence de cet autre point fixe est régit par la pente de  $f$  en 1 c'est à dire  $f'(1) = \mu$ . Voir la figure 3. Ainsi, lorsque  $\mu \leq 1$  il n'y aura pas d'autre point fixe, et si  $\mu > 1$  il y en aura exactement un autre. Donc si  $\mu \leq 1$ , la probabilité d'extinction est 1 mais si  $\mu > 1$  quelle autre point fixe choisir? Et bien 1 est si  $\mu > 1$  un point fixe répulsif tandis que l'autre point fixe est attractif de sorte que la probabilité d'extinction est cet autre point fixe de  $f$ . Finalement nous venons d'établir le résultat suivant.

**Théorème 3.** *La probabilité d'extinction  $q$  est le plus petit point fixe de  $f$  sur  $[0, 1]$ . Si  $\mu \leq 1$ ,  $f$  a un unique point fixe 1, ainsi  $q = 1$  i.e. on a presque sûrement extinction. Si  $\mu > 1$ ,  $f$  a un unique autre point fixe dans  $[0, 1]$  et l'on a donc  $q < 1$ . De plus, dans tous les cas,  $q$  est un point fixe attractif,  $f'(q) < 1$ , et plus précisément pour tout  $s \in [0, 1]$ ,  $f^n(s) \rightarrow q$  lorsque  $n \rightarrow \infty$ .*

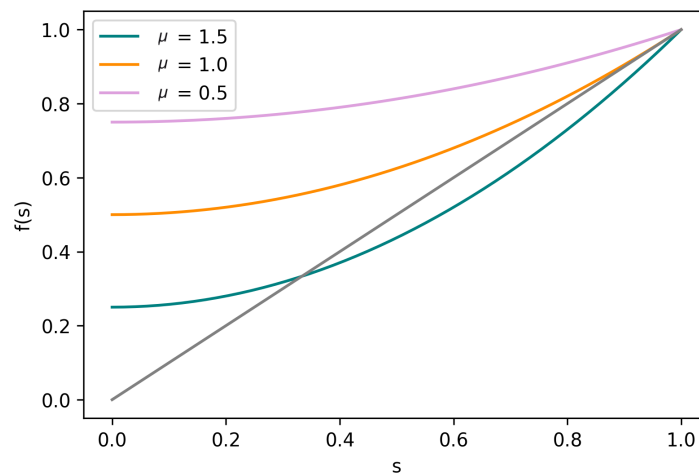


FIGURE 3 – Illustration fonctions génératrices et points fixes. Cas du branchement binaire  $f(s) = (1 - p) + ps^2$  avec  $p = 1/4$  (en violet, cas sous-critique),  $p = 1/2$  (en orange, cas critique) et  $p = 3/4$  (en bleu, cas sur-critique). En gris est représentée la droite  $y = x$ .

**Exemple.** *Loi de reproduction à support dans  $0 - 2$*

Considérons le cas d'une loi de reproduction à support dans  $\{0, 1, 2\}$ . On note  $f(s) = a + bs + cs^2$ , avec afin d'être dans le cas sur-critique. On cherche alors à résoudre l'équation

$$a + (b - 1)s + cs^2 = 0.$$

Les solutions de cette équation sont

$$\frac{1 - b \pm \sqrt{(1 - b)^2 - 4ac}}{2c}.$$

Notons que  $a + b + c = 1$  de sorte que  $(1 - b)^2 - 4ac = (a + b)^2 - 4ac = (a - b)^2$ . Ainsi, la solution non triviale de l'équation est  $\frac{a}{c}$ .

### 1.3 Comportement vers l'explosion

Nous avons vu avec la formule (4) que  $Z_n$  croît en moyenne comme  $\mu^n$ . Mais à quel point la croissance du processus sur l'évènement d'explosion est dicté par  $\mu^n$ . Pour étudier cela, on est amené à introduire le processus normalisé

$$W_n := \frac{Z_n}{\mu^n}.$$

Il se trouve que ce processus est une *martingale*. Étant positive, elle converge dans presque sûrement vers une variable aléatoire  $W$  intégrable donc en particulier presque sûrement finie. Sur les issues  $\omega$  tel que  $W(\omega) \neq 0$ ,  $\mu^n$  est la bonne croissance du processus

$$Z_n(\omega) = \mu^n W_n(\omega) \sim \mu^n W(\omega).$$

Il convient de s'intéresser à l'évènement  $\{W = 0\}$ . Notons qu'il est clair que l'évènement  $W = 0$  contient l'évènement d'extinction. Ce qui nous intéresse est la possibilité que  $W = 0$  sur l'évènement de non-extinction. En fait, on peut voir  $\mathbf{P}(W = 0)$  est un point fixe de  $f$  de sorte que soit  $W = 0$  est l'évènement d'extinction soit  $W = 0$  p.s.

**Proposition 4.** *L'évènement  $W = 0$  est soit égal à l'évènement d'extinction à un ensemble de mesure nulle près, soit est presque certain.*

Le cas où  $W = 0$  p.s. correspond à une perte de masse : il n'y a pas la convergence  $L^1$  de la martingale. On peut en fait montrer que c'est une équivalence. En particulier, ce cas est exclu lorsque  $f$  admet un moment d'ordre 2 car alors la martingale  $W_n$  est uniformément borné dans  $L^2$  donc converge dans  $L^2$  donc également dans  $L^1$ . En fait, une hypothèse plus faible suffit et caractérise la dichotomie. C'est un résultat de KESTEN et SYGMUND datant de 1966.

**Théorème 5.** *La variable aléatoire  $W$  est non nulle presque sûrement ssi  $\mathbf{E}[X \log(X)] < \infty$ .*

On peut interpréter ce résultat ainsi : si la queue de la loi de reproduction ne décroît pas assez vite, c'est à dire si des familles nombreuses surviennent avec une probabilité trop grande, alors comme la moyenne tient davantage compte des grands évènements, la taille typique  $t$  des familles est inférieure à cette dernière, puis  $Z_n$  croît typiquement comme  $t^n = o(\mu^n)$ . Ainsi, la moyenne  $\mu$  n'est pas le bon taux de croissance du modèle.

Le fait remarquable est que lorsque le critère précédent est satisfait, il y a croissance uniforme du processus sur l'évènement d'explosion. Il ne peut apparaître différents taux de croissances. C'est en fait toujours le cas lorsque la condition précédente échoue. C'est un résultat de SENETA (1968) et HEYDE (1970) qui dit qu'il existe une suite  $c_n$  tel que  $c_n^{-1} Z_n$  converge p.s. vers une variable aléatoire non dégénérée, nulle seulement sur l'évènement d'extinction.

## 1.4 Comportement vers l'extinction

Passons maintenant au comportement vers l'extinction. Pour étudier cela, les mathématiciens se sont intéressés au processus

$$Y_n := Z_n \mid n < T < \infty$$

où  $T = \inf\{n : Z_n = 0\}$  est le temps d'extinction. Notamment, YAGLOM (1947) qui montra la convergence du processus.

**Théorème 6.** *Soit  $Z_n$  un processus de Galton-Watson. Le processus  $Y_n$  converge vers une distribution de probabilité  $\pi$  entièrement caractérisé par l'équation*

$$\pi \circ f(s) = f'(q) \cdot \pi(s) + (1 - f'(q)). \quad (6)$$

*C'est à dire que  $\pi(s)$  est l'unique solution de (6) parmi les lois de probabilités sur  $\mathbf{N}^*$ .*

Dans le cas critique,  $Y_n$  converge également en loi mais la limite précédente est triviale. Cela illustre que le processus  $Y_n$  s'en va à l'infini. Pour obtenir une limite non triviale, il faut normaliser le processus afin d'empêcher sa fuite. Pour obtenir la bonne normalisation, le problème consiste à estimer

$$\mathbf{E}(Y_n) = \mathbf{E}[Z_n \mid Z_n > 0].$$



Remarquons que

$$\mathbf{E}[Z_n | Z_n > 0] = \frac{\mathbf{E}[Z_n \mathbf{1}_{Z_n > 0}]}{\mathbf{P}(Z_n > 0)} = \frac{1}{\mathbf{P}(Z_n > 0)}.$$

Le problème est donc équivalent à estimer  $\mathbf{P}(Z_n > 0)$ . En 1938, KOLMOGOROV établit le résultat suivant.

**Théorème 7.** *Pour  $Z_n$  un processus de Galton-Watson critique de variance  $\sigma^2 < \infty$ ,*

$$\mathbf{E}[Z_n | Z_n > 0] = \frac{1}{\mathbf{P}(Z_n > 0)} \underset{n \rightarrow \infty}{\sim} \frac{n\sigma^2}{2}. \quad (7)$$

Ainsi, dans le cas critique on est amené à étudier le processus

$$\hat{Y}_n := \frac{1}{n} Z_n | Z_n > 0.$$

C'est encore YAGLOM (1947) qui sous l'hypothèse d'un moment d'ordre 2 montra la convergence du processus. Cette fois si la limite est explicite mais surtout universelle ce qui fait tout l'intérêt du résultat.

**Théorème 8.** *Sous l'hypothèse que la loi de reproduction  $f$  admet un moment d'ordre 2,*

$$\hat{Y}_n \xrightarrow{\text{Loi}} \mathbf{E}\left(\frac{2}{\sigma^2}\right).$$

## Annexe - Arbre de Galton-Watson

Dans cette annexe, on construit une mesure de probabilité sur les arbres afin que  $Z_n$ , le nombre de sommets à la profondeur  $n$ , définit un processus de Galton-Watson standard. On commence par définir une représentation formelle des arbres, la notation de Neveu, qui consiste à numéroté les sommets d'un arbre via la suite des positions des ancêtres dans leur fraterie. La figure 4 donne un exemple d'arbre et sa représentation de Neveu associée.

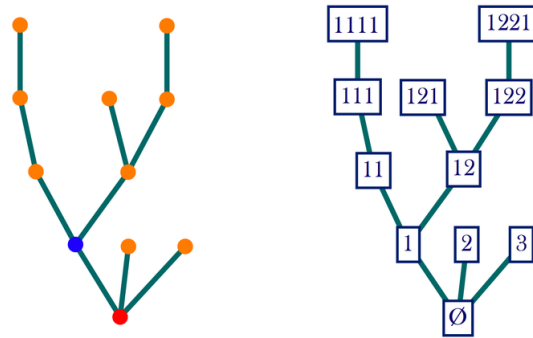


FIGURE 4 – Un arbre et la notation de Neveu associé.

Un arbre sera formellement défini comme l'ensemble des numérotations de Neveu de ses sommets. On désigne par

$$\mathcal{U} = \bigcup_{n=0}^{\infty} (\mathbf{N}_0)^n$$

l'ensemble des étiquettes, c'est à dire des suites finies d'entiers strictement positifs, où par convention  $(\mathbf{N}_0)^0 = \emptyset$ . Si  $u = (u_1, \dots, u_i)$  et  $v = (v_1, \dots, v_j)$  sont deux éléments de  $\mathcal{U}$ , on désigne par  $uv$  la concaténation de  $u$  et  $v$  c'est à dire  $uv = (u_1, \dots, u_i, v_1, \dots, v_j)$ .

**Définition 2.** *Un arbre  $\tau$  est une partie de  $\mathcal{U}$  tel que :*

- i.  $\emptyset \in \tau$ ,

- ii.  $uv \in t \Rightarrow u \in t$ ,
- iii.  $u \in t \Rightarrow \exists k_u(t)$  tel que pour chaque  $j \in \mathbf{N}_0$ ,  $uj \in t$  si et ssi  $1 \leq j \leq k_u(t)$ .

Les éléments de  $\tau$  correspondent au sommet de l'arbre. La profondeur d'un sommet  $u$  est la longueur de la suite  $u$  noté  $|u|$ . L'élément  $\emptyset$  correspond à l'ancêtre commun, c'est l'unique sommet à la profondeur 0. On note  $z_n(\tau)$  l'ensemble des sommets à la profondeur  $n$  c'est à dire

$$z_n(\tau) = \{u \in \tau : |u| = n\}$$

puis  $Z_n(\tau)$  le cardinal de  $z_n(\tau)$  i.e.  $Z_n(\tau) = |z_n(\tau)|$ . Enfin, on désignera par  $\mathcal{T}$  l'ensemble des arbres. La construction de la mesure de Galton-Watson sur  $\mathcal{T}$  repose sur l'observation fondamentale suivante :

**Proposition 9.** *Soit  $Z_n$  un processus de Galton-Watson de loi  $f$ . Conditionnellement à  $Z_n = k$  le processus  $(Z_{n+m} : m \geq 0)$  à la loi de la somme de  $k$  processus de Galton-Watson de loi  $f$ .*

Cette propriété est appelée *propriété de branchement*. En reformulant cette propriété en terme d'arbre il apparait qu'un arbre de Galton-Watson conditionnée à avoir  $k$  sommets à la une génération  $n$ , alors les arbres engendrés par ses  $k$  sommets sont à leur tour des arbres de Galton-Watson. Nous désignerons pour un sommet  $u \in \tau$ ,  $\theta_u \circ \tau$  l'arbre engendré le sommet  $u$ . Cette propriété est la propriété fondamentale des arbres de Galton-Watson. Avec la donnée de la loi de reproduction, elle caractérise entière les arbres de Galton-Watson. C'est intuitivement clair, les arbres sont construits par un processus récursif. Néanmoins la démonstration est technique, et fait intervenir des problèmes technique de théorie des mesures.

**Proposition 10.** *Il existe une unique mesure de probabilité  $\text{GW}_f$  sur  $\mathcal{T}$  tel que :*

1.  $Z_1$  à la loi de  $f$  ;
2. Conditionnellement à  $Z_1$ , les variables aléatoires  $\theta_u$  pour  $u \in z_1$  sont indépendantes de loi GW.

*Preuve.* Si une mesure de probabilité  $\rho$  sur  $\mathcal{T}$  satisfait la propriété de branchement, et que  $Z_1$  à la loi de  $f$  alors

$$\rho(\tau) = \rho(Z_1 = k(\tau) \text{ et } \theta_u \circ \tau = \tau_u, \forall u \in z_1(\tau)) = f[k(\tau)] \prod_{\substack{u \in \tau \\ |u|=1}} \rho(\tau_u)$$

où  $\theta_u$  est la fonction qui à un arbre envoie l'arbre engendré par le sommet  $u$ . Ainsi, en désignant par  $\rho^n$  la marginale sur  $\mathcal{F}_n = \sigma(k_u : |u| \leq n)$ , c'est à dire

$$\rho^n(\tau) = \rho(\tau' : \tau \stackrel{n}{=} \tau')$$

où  $\tau \stackrel{n}{=} \tau'$  signifie que les  $n$  premières générations des deux arbres coïncident. On a alors,

$$\rho^n(\tau) = f[k(\tau)] \prod_{\substack{u \in \tau \\ |u|=1}} \rho^{n-1}(\tau_i).$$

D'où en itérant

$$\rho^n(\tau) = \prod_{\substack{u \in \tau \\ |u| \leq n}} f[k_u(\tau)]. \quad (*)$$

En partant de (\*), il est possible de construire récursivement la mesure de Galton-Watson. On définit pour cela, les mesures  $\rho^n$  selon (\*). Ainsi définit,  $\rho^n$  est une mesure de probabilité sur  $\mathcal{F}_n$ . Les  $\mathcal{F}_n$  forment une *filtration* sur  $\mathcal{T}$  - une suite croissante de tribus et ces mesures sont *compatibles* pour cette filtration au sens où  $\rho^{n+1}$  restreint à  $\mathcal{F}_n$  coïncide avec  $\rho^n$ . Le *théorème d'extension de Kolmogorov* garantit alors l'existence d'une mesure  $\rho$  sur  $\mathcal{F} = \bigcup_{n \in \mathbf{N}} \mathcal{F}_n$  prolongeant les  $\rho^n$ . On peut voir la tribu  $\mathcal{F}$  comme la tribu engendrée par les événements  $\mathcal{T}_u := \{\tau : u \in \tau\}$ .  $\square$

Finalement, vérifions que sous  $\text{GW}_f$ , la suite  $(Z_n : n \geq 0)$  est un processus de Galton-Watson standard de loi  $f$ .

**Proposition 11.** *Sous  $\text{GW}$ ,  $(Z_n : n \geq 0)$  est un processus de Galton-Watson standard de loi de reproduction  $f$ .*

## 2 L'arbre biaisé

Dans cette section, on présente un concept développé par R.LYONS, R.PEMENTLE et Y.PEREZ dans l'article [LPP95] : l'arbre biaisé. Cette notion permet la démonstration de la plupart des résultats mentionnés précédemment. Parfois plus simple, parfois plus compliqués, ces preuves restent dans tous les cas plus éclairantes que d'autres preuves plus élémentaires et calculatoires. La notion introduite par les trois auteurs est le *processus de Galton-Watson biaisé*. C'est un processus de Galton-Watson standard où chaque variable aléatoire a été biaisé au sens suivant.

**Définition 3.** Si  $Z$  est une variable aléatoire avec une moyenne  $\mu$  à valeurs dans  $\mathbf{N}$ , on dit qu'une variable aléatoire  $Z^b$  à la loi du biais de  $Z$  lorsque

$$\mathbf{P}(Z^b = k) = \frac{k\mathbf{P}(Z = k)}{\mu}. \quad (8)$$

Dans ce cas, pour tout fonction  $f : \mathbf{Z} \rightarrow \mathbf{R}$  borné,

$$\mathbf{E}(f(Z^b)) = \frac{1}{\mu} \mathbf{E}(Zf(Z)). \quad (9)$$

A cette époque, le concept d'*arbre biaisé* n'est pas une nouveauté. Elle apparait dans de nombreux travaux à partir du milieu des années 1975. L'avancée de l'article est l'application de ce type de construction au processus de Galton-Watson qui leur permet de redécouvrir des résultats notables de la théorie. Les auteurs proposent en plus une construction du processus biaisée suivant le formalisme des arbres. Cette construction est la suivante :

- On part d'un ancêtre commun  $v_0$  auquel on attribue un nombre aléatoire d'enfants suivant la loi du biais de  $f$ ,
- On choisit parmi ces enfants, uniformément, un enfant  $v_1$ . Cette individu  $v_1$  aura des enfants selon la loi du biais de  $f$  tandis que les autres auront des enfants selon  $f$ .
- On choisit ensuite parmi les enfants de  $v_1$ , un individu  $v_2$  et on applique le point précédent et ainsi de suite.

Au final on a une suite  $(v_i)$  de descendants appelée *épine dorsale* où tous les sommets ont une loi de reproduction biaisée, et sur les enfants n'appartenant pas à l'épine se greffent des processus de Galton-Watson simples.

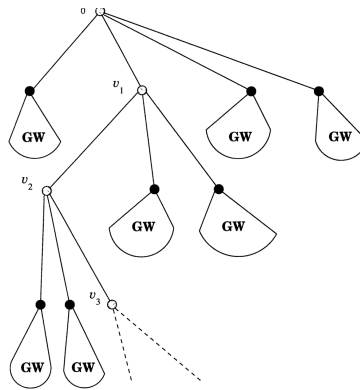


FIGURE 5 – Illustration construction arbre biaisée. Image extraite de [LPP95] figure 1.

Nous laissons en annexe les détails mathématiques. Le fait est qu'on peut définir une mesure  $\mathbf{GW}_f^b$  sur l'espace des arbres telle que sous  $\mathbf{GW}_f^b$ ,  $(Z_n | n \geq 0)$  soit un processus ayant la loi du biais d'un processus de Galton-Watson de loi  $f$ .

## 2.1 Immigration

Le processus de Galton-Watson biaisé, peut se voir comme un processus de Galton-Watson avec *immigration* c'est à dire où à chaque génération  $n$ , arrive de nouveaux individus dits *immigrés* quantifiés par une variable aléatoire  $I_n$ . Les variables aléatoires  $I_n$  sont indépendantes et seront prises identiquement distribuées.

**Définition 4.** Le processus de Galton-Watson de loi de reproduction  $f$  avec immigration  $I$  est le processus  $Z_n^I = Z_n$  définie par  $Z_0 = I_0$  puis,

$$Z_{n+1} = I_{n+1} + \sum_{i=1}^{Z_n} X_{i,n} \quad (10)$$

où les variables aléatoires  $X_{i,n}$  sont i.i.d de loi  $f$  et les variables aléatoires  $I_n$  i.i.d distribué comme  $I$ .

Si l'on note  $\mathcal{I}$  la tribu d'immigration engendrée par les variables aléatoires  $I_n$ , on voit que

$$Z_n^I | \mathcal{I} \stackrel{\text{Loi}}{=} \sum_{k=0}^n Z_{n-k}^{(k)} \quad (11)$$

où les  $Z^{(k)}$  sont des processus de Galton-Watson indépendant de loi  $f$  et de population initiale  $I_k$ . Ainsi, un processus de Galton-Watson avec immigration est simplement un processus tel qu'à chaque génération s'ajoute un nouveau processus de Galton-Watson. En particulier, on a une expression de la fonction génératrice du processus.

$$\mathbf{E}[s^{Z_n^I}] = s \prod_{k=0}^{n-1} g \circ f^k(s) \quad (12)$$

où  $g$  désigne la fonction génératrice de  $I$ .

**Proposition 12.** Le processus de Galton-Watson biaisée à la loi d'un processus de Galton-Watson avec immigration de loi  $g[k+1] = f^b[k]$  et  $g[0] = 0$  avec  $f^b$  la loi du biais de  $f$ .

## 2.2 Critère $X \log(X)$

On illustre l'utilisation de l'arbre biaisé avec une preuve alternative du théorème de Kesten-Sigmund (théorème 5) décrite dans [LPP95]. Pour rappel, on se place dans le régime sur-critique et l'on cherche à montrer la dichotomie suivante :

$$W = 0 \text{ p.s} \Leftrightarrow \mathbf{E}(W) = 0 \Leftrightarrow \mathbf{E}(X \log(X)) = \infty, \quad (13)$$

$$W \neq 0 \text{ p.s} \Leftrightarrow \mathbf{E}(W) = 1 \Leftrightarrow \mathbf{E}(X \log(X)) < \infty. \quad (14)$$

Le processus biaisé est bien adapté à l'étude de la martingale  $W_n$  d'après la formule (31). On aimerait passer cette égalité à la limite  $n \rightarrow \infty$ . Avant de faire cela, il faut prendre des précautions. D'abord,  $W = \lim W_n$  n'est définie que GW-presque partout, pour pouvoir parler de  $W$  sous n'importe quelle mesure on pose  $W = \limsup W_n$ . Alors, on peut montrer que pour tout  $A \in \mathcal{F}$ ,

$$\text{GW}^b(A) = \mathbf{E}(\mathbb{1}_A W) + \text{GW}^b(A \cap \{W = \infty\}) \quad (15)$$

De sorte que la dichotomie (13)-(14) devient

$$W = 0 \text{ GW-p.s.} \Leftrightarrow W = \infty \text{ GW}^b\text{-p.s.}, \quad (16)$$

$$\mathbf{E}(W) = 1 \Leftrightarrow W < \infty \text{ GW}^b\text{-p.s.} \quad (17)$$

Ainsi, nous avons considérablement simplifier le problème et le théorème 5 découle du théorème suivant dû à Seneta (1970) sur les processus de Galton-Watson avec immigration, qui est nettement plus simple à aborder. Pour la démonstration, on pourra se référer à [LPP95].

**Théorème 13.** Soit  $Z_n^I$  un processus de Galton-Watson de moyenne  $\mu > 1$  avec une immigration  $I$ . Si  $\mathbf{E}[\log I] < \infty$  alors  $\mu^{-n} Z_n$  converge presque sûrement vers une limite fini et sinon,  $\limsup_{n \rightarrow \infty} c^{-n} Z_n = \infty$  p.s pour toute constante  $c > 0$ .

*Preuve de la formule (15).* Soit  $A \in \mathcal{F}$ . Pour pouvoir utiliser la formule (31), il faut se ramener à une marginale. Pour cela on considère

$$A_n = \{t : \exists t' \in A \text{ t.q. } t \stackrel{n}{=} t'\}.$$

Ainsi,  $A_n \in \mathcal{F}_n$  définit une suite décroissante de limite  $A$ . Nous avons alors,

$$\text{GW}^b(A) = \lim_{n \rightarrow \infty} \mathbf{E}(W_n \mathbf{1}_{A_n}).$$

Pour pouvoir passer à la limite sous l'espérance, il nous faut une hypothèse de domination. Ce n'est pas émergent dans notre situation, nous allons la forcer en intersectant avec l'évènement  $\{\sup_{k \leq n} W_k \leq u\}$ . Cette famille d'évènements est croissante en  $u$ , de limite  $\{W < \infty\}$ . Nous avons alors,

$$\text{GW}^b(A \cap \{W < \infty\}) = \lim_{u \rightarrow \infty} \text{GW}^b(A \mathbf{1}_{\sup_{k \leq n} W_k \leq u}) = \lim_{u \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbf{E}(W_n \mathbf{1}_{A_n} \mathbf{1}_{\sup_{k \leq n} W_k \leq u}).$$

Maintenant, on peut passer à la limite  $n \rightarrow \infty$  dans l'espérance

$$\lim_{n \rightarrow \infty} \mathbf{E}(W_n \mathbf{1}_{A_n} \mathbf{1}_{\sup_{k \leq n} W_k \leq u}) = \mathbf{E}(W \mathbf{1}_A \mathbf{1}_{\sup_{k \in \mathbb{N}} W_k \leq u}).$$

Finalement, on peut passer à la limite  $u \rightarrow \infty$  dans le terme de droite par convergence monotone pour obtenir

$$\text{GW}^b(A \cap \{W < \infty\}) = \mathbf{E}(W \mathbf{1}_A).$$

□

### 2.3 Théorème de convergence de Yaglom critique

Le processus biaisé permet également d'étudier le processus de Galton-Watson critique conditionnel  $Y_n = Z_n \mid Z_n > 0$ . En effet, si l'on impose l'épine d'un processus de Galton-Watson biaisé, on retrouve le processus  $Y_n$ . Pour simplifier les choses, on fait le choix arbitraire de fixer l'épine tout à gauche. On considère ainsi  $T^b$  un arbre de Galton-Watson biaisé dont on note  $V$  l'épine. On notera  $Z_n^b$  la variable aléatoire  $Z_n(T^b)$ . On commence par décomposer  $Z_n^b$  suivant les individus à gauche de l'épine et les individus à droite  $Z_n = L_n + R_n$  où  $L_n$  (resp.  $R_n$ ) sont les sommets de la  $n$ -ème génération à gauche (resp. droite) de  $v_n$  exclu (resp. inclu). On affirme alors qu'en désignant par  $A_n$  l'évènement  $\{L_n = 0\}$  on a,

$$Z_n^b \mid A_n \stackrel{\text{Loi}}{=} Z_n \mid Z_n > 0 \quad (18)$$

On voit rapidement cela en observant que  $\mathbf{P}(L_n = 0 \mid Z_n^b) = 1/Z_n^b$  puis en inversant le conditionnement avec la formule de Bayes.

Nous allons utiliser ce point de vue pour retrouver les résultats de KOLMOGOROV (théorème 7) dans un premier temps et de YAGLOM (théorème 8) dans un second temps.

#### Estimation de Kolmogorov

On écrit la décomposition canonique d'un processus avec immigration  $Z_n^b = 1 + \sum_{j=1}^n Z_{n,j}^b$  où  $Z_{n,j}^b$  correspond aux individus de la génération  $n$  descendant de  $V_{j-1}$  mais pas de  $V_j$ . D'après la proposition 12,  $Z_{n,j}^b$  est un processus de Galton-Watson critique de population initiale  $I_j \stackrel{\text{Loi}}{=} X^b - 1$ . On décompose de la même façon  $L_n$  et  $R_n$  en

$$L_n = \sum_{j=1}^n L_{n,j}, \quad R_n = 1 + \sum_{j=1}^n R_{n,j}$$

où  $Z_{n,j}^b = L_{n,j} + R_{n,j}$  avec  $L_{n,j}$  (resp.  $R_{n,j}$ ) le nombre de descendants des immigrés de la génération  $j$  à gauche (resp. droite) de  $v_n$  exclu.<sup>5</sup> Alors  $A_n = \bigcap_{j=1}^n A_{n,j}$  où  $A_{n,j}$  est l'évènement  $L_{n,j} = 0$ .

On commence donc par s'intéresser à l'estimation de Kolmogorov 7. Nous avons avec le formalisme précédent

$$\mathbf{E}[Z_n | Z_n > 0] = \mathbf{E}[Z_n^b | A_n] = 1 + \sum_{j=1}^n \mathbf{E}[Z_{n,j} | A_{n,j}].$$

L'intuition est alors la suivante, l'espérance de  $Z_{n,j}$  est  $\sigma^2$ , on s'attend à ce que  $\frac{\sigma^2}{2}$  de ces sommets se trouvent à gauche de  $v_n$  et  $\frac{\sigma^2}{2}$  à droite. En conditionnant par  $A_{n,j}$ , il ne reste plus que les sommets à droite c'est à dire  $\frac{\sigma^2}{2}$ . Partant de cette intuition, il nous faut comparer les variables aléatoires  $Z_{n,j} | A_{n,j}$  et  $R_{n,j}$ . C'est ce que fait l'article en construisant une variable aléatoire  $\tilde{R}_n$  "proche" de  $R_n$  et de loi  $Z_n^b | A_n$ . Pour réaliser cela ils posent  $\tilde{R}_n = 1 + \sum_{j=1}^n \tilde{R}_{n,j}$  où

$$\tilde{R}_{n,j} = R_{n,j} \mathbf{1}_{A_{n,j}} + R'_{n,j} \mathbf{1}_{A_{n,j}^c}$$

avec  $R'_{n,j}$  est une variable aléatoire indépendante de toute autre variable aléatoire précédemment introduite ayant la loi de  $R_{n,j} | A_{n,j}$ . Effectivement, nous avons les égalités de lois

$$Z_n^* | A_n = R_n | A_n \stackrel{\text{Loi}}{=} 1 + \sum_{j=1}^n R_{n,j} | A_{n,j} \quad (19)$$

car les variables aléatoires  $(L_{n,j}, R_{n,j})$  pour  $1 \leq j \leq n$  sont indépendante. Et un rapide calcul donne

$$\mathbf{P}(R_{n,j} | A_{n,j}) = \mathbf{P}(R_{n,j} \text{ et } A_{n,j}) + \mathbf{P}(R_{n,j} | A_{n,j}) \mathbf{P}(A_{n,j}^c) = \mathbf{P}(R_{n,j} \text{ et } A_{n,j}) + \mathbf{P}(R'_{n,j} \text{ et } A_{n,j}^c). \quad (20)$$

Ainsi,  $\tilde{R}_n$  est une variable aléatoire de loi  $Z_n^b | A_n$ . Nous allons maintenant comparer  $R_n$  et  $\tilde{R}_n$  en distance moyenne. Du fait de notre construction, nous avons l'inégalité suivante

$$\mathbf{E}(|R_n - \tilde{R}_n|) \leq \sum_{j=1}^n \mathbf{E}[\mathbf{1}_{A_{n,j}^c} |R_{n,j} - R'_{n,j}|] \quad (21)$$

Il ne reste plus qu'à estimer nos variables aléatoires sur les petits bouts  $A_{n,j}^c$ . Ne disposant pas d'information supplémentaire sur ces évènements, on majore brutalement

$$\sum_{j=1}^n \mathbf{E}[\mathbf{1}_{A_{n,j}^c} |R_{n,j} - R'_{n,j}|] \leq \sum_{j=1}^n \left( \mathbf{E}[\mathbf{1}_{A_{n,j}^c} R_{n,j}] + \mathbf{E}[\mathbf{1}_{A_{n,j}^c} R'_{n,j}] \right). \quad (22)$$

Commençons par traiter le premier terme. Pour cela notons  $N_j$  l'entier tel que  $V_{j-1}N_j = V_j$  c'est à dire la position de  $V_j$  dans les enfants de  $V_{j-1}$ . En fixant  $I_j$  et  $N_j$  les variables aléatoires  $R_{n,j}$  et  $L_{n,j}$  deviennent indépendantes d'où

$$\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c} | I_j, N_j] = \mathbf{E}[R_{n,j} | I_j, N_j] \mathbf{E}[\mathbf{1}_{A_{n,j}^c} | I_j, N_j]. \quad (23)$$

Or  $Z_{n,j}^b$  étant un **GW**-critique de population initiale  $I_j$ ,

$$\mathbf{E}[R_{n,j} | I_j, N_j] \leq \mathbf{E}[Z_{n,j}^b | I_j, N_j] = I_j \quad (24)$$

Donc en réinjectant dans (23),

$$\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \mathbf{E} \left[ I_j \mathbf{P}(A_{n,j}^c | I_j, N_j) \right]. \quad (25)$$

Maintenant  $\mathbf{1}_{A_{n,j}^c} \leq \mathbf{1}_{Z_{n,j}^b > 0}$  et cette dernière variable aléatoire converge presque sûrement vers 0 car  $Z_{n,j}$  est un **GW**-critique. Sachant que  $I_j$  est intégrable on a par convergence dominée,  $\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] \rightarrow$

5. Le 1 dans l'écriture de  $R_n$  correspond justement à  $v_n$ .

0. Passons au second terme. Tout d'abord les variables aléatoires  $R'_{n,j}$  et  $\mathbf{1}_{A_{n,j}^c}$  sont indépendantes d'où

$$\mathbf{E}[\mathbf{1}_{A_{n,j}^c} R'_{n,j}] = \mathbf{E}(R'_{n,j})\mathbf{P}(A_{n,j}^c) = \mathbf{E}(Z_{n,j}^b | A_{n,j})\mathbf{P}(A_{n,j}^c). \quad (26)$$

Maintenant, les variables aléatoires  $Z_{n,j}^b$  et  $\mathbf{1}_{A_{n,j}}$  sont négativement corrélés donc

$$\mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \sigma^2 \mathbf{P}(A_{n,j}^c) \quad (27)$$

Mais comme nous l'avons montré précédemment, le terme de droite tend vers 0 lorsque  $n \rightarrow \infty$ . Finalement,

$$\mathbf{E}(|R_n - \tilde{R}_n|) \underset{n \rightarrow \infty}{=} o(n). \quad (28)$$

Nous en déduisons l'approximation de Kolmogorov (théorème 7),

$$\frac{1}{n} \mathbf{E}(Z_n | Z_n > 0) = \frac{1}{n} \mathbf{E}(\tilde{R}_n) = \frac{1}{n} \mathbf{E}(R_n) + o(1) \longrightarrow \frac{\sigma^2}{2}. \quad (29)$$

### Limite de Yaglom

Nous allons pousser le point de vue précédent pour retrouver la convergence du processus conditionnel de  $\frac{1}{n}Z_n | Z_n > 0$  vers la loi exponentielle de paramètre  $\frac{2}{\sigma^2}$ . Cette démarche est encore une fois présentée dans [LPP95]. Les auteurs mettent préalablement en lumière la propriété suivante.

**Proposition 14.** *Une variable aléatoire  $Y$  suit une loi exponentielle si et ssi, pour une variable aléatoire uniforme  $U$  indépendante de  $Y^b$ ,  $U \cdot Y^b$  à la loi de  $Y$ .*

L'intuition est alors la suivante. En fixant l'épine tout à gauche,  $Z_n^b$  se réduit aux individus à droite c'est à dire  $R_n$ . Or,

$$\frac{1}{n} R_n = \frac{1}{n} [U \cdot Z_n^b] \underset{n \rightarrow \infty}{\sim} \frac{1}{n} U \cdot Z_n^b$$

où  $U$  est une variable aléatoire uniforme sur  $[0, 1]$  indépendante de  $Z_n^b$ . Encore une fois, le problème est de comparer les variables aléatoires  $Z_n^b | A_n$  et  $R_n$ . On réutilise naturellement la variable aléatoire  $\tilde{R}_n$ . Au vu de l'approximation de Kolmogorov, la suite  $\frac{1}{n} \tilde{R}_n$  est uniformément bornée dans  $L^1$  donc est tendue. Il s'agit donc de montrer que la suite converge en distribution vers une variable aléatoire  $Y$  alors,  $Y$  suit une loi exponentielle.<sup>6</sup> Le même argument s'applique à  $Z_n/n$ , et quitte à regarder une suite extraite, on peut supposer que  $Z_n^b/n$  converge vers une variable aléatoire  $Z$ . Comme, la première suite est biaisée par rapport à la deuxième, on voit rapidement que  $Y$  biaisée par rapport à  $Z$ . Selon, la proposition 14, il faut voir que  $Y$  à la loi de  $U \cdot Z$  pour une variable aléatoire uniforme indépendante de  $Z$ . Mais, d'après (28),  $\frac{1}{n} R_n$  converge également en distribution vers  $Y$ , et comme nous l'avons dit précédemment,  $R_n$  à la loi de  $[U \cdot Z_n^b]$  pour une variable uniforme  $U$  indépendante de toutes les autres. Or,

$$\left| \frac{1}{n} U \cdot Z_n^b - \frac{1}{n} [U \cdot Z_n^b] \right| \leq \frac{1}{n} \quad (30)$$

donc ces deux variables aléatoires converge en distribution vers la même limite qui est respectivement  $U \cdot Z$  et  $Y$ . Et voilà qui conclut le raisonnement.

### Annexe - Arbre de Galton-Watson biaisé

Dans cette annexe, nous allons voir comment définir la mesure de Galton-Watson biaisé à partir de la recette de [LPP95] que nous rappelons :

- On part d'un ancêtre commun  $v_0$  auquel on attribue un nombre aléatoire d'enfants suivant la loi du biais de  $f$ ,

---

6. Voir annexe convergences de mesures



- On choisit parmi ces enfants, uniformément, un enfant  $v_1$ . Cette individu  $v_1$  aura des enfants selon la loi du biais de  $f$  tandis que les autres auront des enfants selon  $f$ .
- On choisit ensuite parmi les enfants de  $v_1$ , un individu  $v_2$  et on applique le point précédent et ainsi de suite.

Commençons par mettre le formalisme nécessaire.

**Définition 5.** Une épine dorsal d'un arbre  $t$  est une suite de sommets  $(v_i)_{0 \leq i \leq \ell(t)}$  partant de l'ancêtre jusqu'à la dernière génération de l'arbre tel que le sommet  $v_{i+1}$  soit un enfant de  $v_i$ . Généralement, nous désignerons par  $\dot{t}$  un arbre avec épine. Nous noterons alors  $t$  l'arbre sous jacent. Ainsi, nous noterons  $\mathcal{T}$  l'ensemble des arbres avec épines.

Formellement, la recette indique comment construire un arbre aléatoire avec épine  $\dot{T}$ . La notice précédente permet d'induire la loi  $\text{GW}^b$  de  $\dot{T}$ . Plus précisément, l'information que nous pouvons récupérer est la loi des marginales. On note  $\text{GW}_n^b$  la marginale sur les  $n$  premières générations c'est à dire sur la tribu la loi induite sur la tribu  $\mathcal{F}_n = \{(t, v) : t \in \mathcal{T}_n\}$ . En notant  $k$  le nombre de sommets de  $\dot{t}$  à la première génération et en désignant par  $i$  l'indice du sommet appartenant à l'épine nous avons l'égalité suivante

$$\text{GW}_{n+1}^b(\dot{t}) = \frac{k f[k]}{\mu} \frac{1}{k} \text{GW}_n^b(\dot{t}_i) \prod_{j \neq i} \text{GW}_n(y_j)$$

où le premier terme  $\frac{k f[k]}{\mu}$  correspond à la probabilité d'avoir  $k$  enfants sous la loi du biais donc que  $Z_1(\dot{T}) = k$ ,  $\frac{1}{k}$  désigne la probabilité que l'indice de sommet de  $\dot{T}$  appartenant à l'épine soit d'indice  $i$ , et les termes suivant sont les probabilités que les arbres induit coïncident. Seul l'arbre engendré par  $i$  garde la dépendance en l'épine dorsal, car c'est le seul à la contenir. En itérant cette égalité nous obtenons

$$\text{GW}_n^b(\dot{t}) = \frac{1}{m^n} \text{GW}_n(t).$$

Nous en déduisons la loi des marginales de  $T$  toujours notés  $\text{GW}_n^b$ ,

$$\text{GW}_n^b(t) = \frac{Z_n(t)}{m^n} \text{GW}_n(t)$$

étant donnée qu'il y a exactement  $Z_n(t)$  façon de munir  $t$  d'une épine dorsal; une épine étant caractérisée par son dernier sommet, l'épine complète étant formé des ancêtres de ce dernier sommet. Qu'avons nous fait jusqu'alors? Nous avons induit d'une notice de construction d'une variable aléatoire, la loi des marginales de cette dernière. Insistons que tout ce raisonnement n'est pas mathématique mais plutôt "méta-mathématique". Du point de vue mathématique, nous allons partir de la dernière égalité. Nous définissons ainsi la suite de mesures  $\text{GW}_n^b$  par

$$\text{GW}_n^b := W_n \cdot \text{GW}_n.$$

Ces équations sont compatibles au sens où pour  $A \in \mathcal{F}_n$ ,

$$\text{GW}_{n+1}^b(A) = \mathbf{E}(W_{n+1} \mathbf{1}_A) = \mathbf{E}(\mathbf{E}(W_{n+1} \mathbf{1}_A \mid \mathcal{F}_n)) = \mathbf{E}(\mathbf{1}_A \mathbf{E}(W_{n+1} \mid \mathcal{F}_n)) = \mathbf{E}(\mathbf{1}_A W_n) = \text{GW}_n^b(A)$$

où les espérances sont prises selon  $\text{GW}$ . Il découle alors du théorème d'extension de Kolmogorov l'existence d'une unique mesure  $\text{GW}^b$  sur  $(\mathcal{T}, \mathcal{F})$  prolongeant les mesures  $\text{GW}_n$ .

**Proposition 15.** Il existe une unique mesure  $\text{GW}^b = \text{GW}_f^b$  sur l'espace des arbres tel que  $Z_1$  ait la loi de  $f$  et,

$$\text{GW}_{n,f}^b = W_n \cdot \text{GW}_{n,f}. \quad (31)$$

Finalement, on peut vérifier que sous  $\text{GW}^b$ , c'est à dire en munissant  $(\mathcal{T}, \mathcal{F})$  de la loi  $\text{GW}^b$ , la suite de variables aléatoires  $Z_n : t \mapsto \#\{u \in t : |u| = n\}$  est un processus de Galton-Watson biaisée :

$$\text{GW}^b(Z_n = k) = \mathbf{E}(W_n \mathbf{1}_{Z_n=k}) = \frac{k}{m^n} \mathbf{E}(\mathbf{1}_{Z_n=k}) = \frac{k}{m^n} \text{GW}(Z_n = k).$$

### 3 Vitesse de convergence vers la limite de Yaglom

Dans cette section, on étudie la vitesse de convergence de  $\hat{Y}_n = \frac{2}{n\sigma^2} Z_n | Z_n > 0$  vers la loi exponentielle standard Exp (théorème 8). La première borne fut obtenue par PEKÖZ et RÖLLIN dans les années 2010 [PR11]. Leur démarche se base sur la *méthode de Stein*, que nous présentons dans la première partie, et réutilise les outils développées par R.LYONS, R.PEMENTLE et Y.PEREZ. Toutefois, une erreur s'est glissée dans leur raisonnement qui fut repérée et corrigée par N.CARDONA-TOBÓN, A.JARAMILLO et S.PALAU dans [CTJP23] avec pour effet de changer légèrement la borne. Mais il me semble que leur correction comporte une inexactitude qui si corrigée, permettrait de retrouver la borne originel.

#### 3.1 Méthode de Stein

La *méthode de Stein* est une technique développée par C. STEIN, à la fin des années soixante, pour obtenir des bornes entre deux lois de probabilités pour certaines. Initialement introduite pour obtenir des bornes avec la loi normale dans le cadre du TCL, elle fut développée et adaptée à d'autres lois de probabilités usuelles. Dans notre cas, nous allons utiliser la méthode de Stein pour obtenir une borne de convergence de  $2Y_n/n\sigma^2$  vers la loi exponentielle standard. L'essentiel de ces deux parties est tiré du tour de vue de R.ROSS sur la méthode de Stein [Ros11]. La méthode s'applique aux distances de la forme

$$d_{\mathcal{H}}(\mu, \nu) = \sup_{h \in \mathcal{H}} \left| \int h d\mu - \int h d\nu \right|$$

où  $\mathcal{H}$  est un ensemble de fonctions tests. Cela comprend :

- la *distance en variation totale* - distance caractérisant la convergence en loi - pour  $\mathcal{H} = \{\mathbf{1}_A : A \text{ borélien de } \mathbf{R}\}$ ;
- la *distance de Kolmogorov* - converge uniforme des fonctions de répartitions - pour  $\mathcal{H} = \{\mathbf{1}_{\cdot \leq x} : x \in \mathbf{R}\}$ ;
- la *distance de Wasserstein* pour  $\mathcal{H} = \{h : \mathbf{R} \rightarrow \mathbf{R} : |h(x) - h(y)| \leq |x - y|\}$ .

La distance de Wasserstein est celle qui se prête le mieux à la méthode de Stein parce que les fonctions utilisées ont de bonnes propriétés. Elle est de plus particulièrement naturelle pour notre problème car la convergence d'une suite  $\mu_n$  de mesures vers une mesure  $\mu$  pour la distance de Wasserstein est équivalente à la convergence en loi ainsi qu'à la convergence du premier moment. Et c'est précisément ce que l'on a d'après les théorèmes 7 et 8. Dans la suite nous travaillerons donc avec la distance de Wasserstein que nous noterons simplement  $d$ . L'ensemble des fonctions tests sera donc

$$\mathcal{H} := \{h : \mathbf{R} \rightarrow \mathbf{R} : |h(x) - h(y)| \leq |x - y|\}$$

c'est à dire l'ensemble des applications 1-Lipschitzienne de  $\mathbf{R}$  dans  $\mathbf{R}$ . Le *théorème de Rademacher* assure qu'une fonction 1-Lipschitzienne  $h : \mathbf{R} \rightarrow \mathbf{R}$  est dérivable presque partout, de dérivée  $h'$  essentiellement bornée et que quelque soit  $(x, y) \in \mathbf{R}^2$ ,

$$h(x) - h(y) = \int_y^x h'(t) dt.$$

Cette égalité rend notamment possible les intégrations par parties :  $\int_a^b g' h = [gh]_a^b - \int_a^b g h'$  pour une fonction lisse  $g$ . Le principe de la méthode de Stein est d'introduire un *opérateur différentiel*  $\mathcal{A}$  tel que pour tout  $h \in \mathcal{H}$ , il existe une solution  $f_h$  au problème suivant

$$\mathcal{A}f = h - \int h d\mu \tag{32}$$

où  $\mu$  est la loi cible. En général, il existe une infinité de solutions au problème (32), on choisit la solution  $f_h$  qui nous arrange. On peut alors réécrire la distance entre  $\mu$  et  $\nu$  comme

$$d(\mu, \nu) = \sup_{h \in \mathcal{H}} \left| \int \mathcal{A} f_h d\mu \right|.$$

A partir de cette expression, on voit que l'opérateur  $\mathcal{A}$  doit être *caractéristique* de la loi  $\nu$  au sens où

$$\int \mathcal{A} f d\mu = 0, \forall f \in \mathcal{F} \text{ ssi } \mu = \nu.$$

Ne connaissant à priori pas  $\mathcal{F}$ , on demande que l'équivalence précédente soit satisfaite pour  $\mathcal{F} = \mathcal{C}_b(\mathbf{R}, \mathbf{R})$ . Un tel opérateur sera appelé *opérateur caractéristique* de  $\nu$ . Si de plus, il existe toujours une solution  $f_h$  à (32) nous dirons que  $\mathcal{A}$  est un *opérateur de Stein* de  $\nu$ . Il n'y a pas d'unicité de l'opérateur de Stein. En pratique, on cherche l'opérateur le plus "simple". Dans notre cas, la loi cible est la loi exponentielle standard Exp. Cherchons à appliquer la méthode de Stein dans ce cas. On commence par chercher un opérateur caractéristique. Pour faire cela, on regarde l'effet d'une dérivation d'une fonction test. Formellement,

$$\int_0^\infty f'(x) e^{-x} dx = [f(x) e^{-x}]_0^\infty + \int_0^\infty f(x) e^{-x} dx = f(0) + \int_0^\infty f(x) e^{-x} dx$$

ce qui suggère de regarder l'opérateur

$$\mathcal{A} f : x \mapsto f'(x) - f(x) + f(0).$$

Il apparaît que  $\mathcal{A}$  est un opérateur caractéristique de Exp. Nous sommes alors amenés à résoudre le problème

$$f' - f + f(0) = h - \int_0^\infty h(t) e^{-t} dt. \quad (33)$$

On montre alors le résultat suivant (voir [Ros11]).

**Proposition 16.** *Pour tout  $h \in \mathcal{H}$ , il existe une unique solution  $f_h$  de (33) s'annulant en 0. Et l'on a les estimations suivantes :*

$$\|f_h'\|_\infty \leq \|h'\|_\infty \quad \|f_h''\|_\infty \leq 2\|h'\|_\infty.$$

Nous pouvons ainsi écrire

$$d_W(Z, \text{Exp}) = \sup_{h \in \mathcal{H}} \mathbf{E}[\mathcal{A} f_h(Z)].$$

Or,  $\mathcal{A} f_h(Z) = f_h'(Z) - f_h(Z)$ . Maintenant, nous allons oublier que  $f_h$  est solution de (32) et simplement retenir les estimations précédentes. Ainsi, nous noterons  $\mathcal{F}$  l'ensemble des fonctions  $f : \mathbf{R} \rightarrow \mathbf{R}$  s'annulant en 0 tel que  $\|f'\| \leq 1$  et  $\|f''\| \leq 2$ . De sorte à écrire,

**Corollaire 17.** *Avec les notations précédentes,*

$$d_W(Y, \text{Exp}) \leq \sup_{f \in \mathcal{F}} |\mathbf{E}[f'(Y) - f(Y)]|. \quad (34)$$

Voilà la base de la méthode de Stein : introduire un opérateur adéquat à notre loi cible afin de transformer notre différence en l'évaluation de l'espérance d'une expression différentielle en notre variable aléatoire. On peut alors traiter ces expressions à l'aide d'outils d'analyse rudimentaires comme les développements de Taylor. Aussi, on peut introduire des objets probabilistes adaptés de sorte à réduire les bornes obtenues à contrôle de l'écart entre différentes variables aléatoires. C'est ce que nous allons faire dans la suite.

Le problème dans l'inégalité (34), c'est que l'on compare une fonction avec sa dérivée, ce qui nous empêche d'utiliser les estimations avec une inégalité des accroissements finis et de se débarrasser de la dépendance en  $f$ . On est ainsi amené à introduire la notion suivante.

**Définition 6.** Soit  $Y$  une variable aléatoire intégrable. On dit qu'une variable aléatoire  $Y^e$  a la distribution de l'équilibre de  $Y$  si pour toute fonction absolument continue  $f$ ,

$$\mathbf{E}[f(Y) - f(0)] = \mu \mathbf{E}[f'(Y^e)].$$

**Corollaire 18.** Soit  $Y$  une variable aléatoire intégrable d'espérance 1. Si  $Y^e$  est une variable aléatoire ayant la distribution de l'équilibre de  $Y$  alors,

$$d_W(Y, \text{Exp}) \leq 2\mathbf{E}[|Y - Y^e|] \quad (35)$$

Le point essentiel est que si  $Y^b$  désigne une variable aléatoire biaisé par rapport à  $Y$  et  $U$  une variable aléatoire uniforme indépendante de  $Y^b$  alors  $U \cdot Y^b$  à la distribution de l'équilibre de  $Y$ . En effet,

$$\mathbf{E}[f'(U \cdot Y^b)] = \mathbf{E}\left(\int_0^1 f'(uY^b)du\right) = \mathbf{E}\left(\frac{f(Y^b)}{Y^b}\right) = \mathbf{E}(f(Y)).$$

### 3.2 Vitesse de convergence

Venons en à notre problème. Nous allons partir du corollaire 18. Tout d'abord, afin de satisfaire les hypothèses, nous devons renormaliser notre processus par la moyenne. Notons  $b_n = \mathbf{E}(Z_n | Z_n > 0)$  et définissons

$$\bar{Y}_n = \frac{1}{b_n} Y_n = \frac{1}{\mathbf{E}(Z_n | Z_n > 0)} Z_n | Z_n > 0.$$

Ensuite nous allons appliquer le corollaire 18 avec  $\tilde{R}_n/b_n$  comme copie de  $\bar{Y}_n$ . Pour  $\bar{Y}_n^e$  commence par remarquer que  $\bar{Y}_n^e \stackrel{\text{Loi}}{=} Y_n^e/b_n$ . Ensuite, voyons que  $Y_n^b \stackrel{\text{Loi}}{=} Z_n^b$ . Finalement, remarquons que  $R_n \stackrel{\text{Loi}}{=} [U \cdot Z_n^b]$  suit une loi uniforme sur  $\{1, \dots, Z_n^b\}$ . Ainsi,

$$Y_n^e \stackrel{\text{Loi}}{=} U \cdot Z_n^b \stackrel{\text{Loi}}{=} R_n - U.$$

Nous prendrons alors  $(R_n - U)/b_n$  comme variable aléatoire ayant la distribution de l'équilibre de  $\bar{Y}_n$ . Finalement nous avons :

$$d_W(\bar{Y}_n, \text{Exp}) \leq \frac{2}{b_n} \left\{ \frac{1}{2} + \mathbf{E}[|\tilde{R}_n - R_n|] \right\}. \quad (36)$$

On est donc ramené à estimer  $\mathbf{E}[|\tilde{R}_n - R_n|]$ , un travail que nous avons déjà entrepris dans la partie 2.3. Dans cette partie, nous n'avions montré que la convergence vers 0 de ce terme. Mais en appliquant la même démarche, et sous l'hypothèse d'un moment d'ordre 3, nous sommes en mesure d'obtenir des bornes effectives qui conduisent au résultat suivant.

**Théorème 19.** On suppose que la loi de reproduction  $f$  admet un moment d'ordre 3. Il existe une constante  $C$  tel que

$$d_W(\bar{Y}_n, \text{Exp}) \leq C \frac{\log(n)}{n}.$$

En pratique il est préférable d'obtenir une borne pour la renormalisation  $\hat{Y}_n = \frac{2}{n\sigma^2} Y_n$ . Rappelons que ce raisonnement est faux avec cette renormalisation car  $\hat{Y}_n$  n'est pas d'espérance 1. C'est en fait l'erreur que font les deux auteurs dans leur article [PR11]. Dans leur article [CTJP23], les auteurs repèrent cette erreur. Pour la corriger, ils reviennent à la borne précédente en écrivant

$$d_W(\hat{Y}_n, \text{Exp}) \leq d_W(\bar{Y}_n, \text{Exp}) + d_W(\bar{Y}_n, \hat{Y}_n). \quad (37)$$

Le deuxième terme se calcule simplement

$$d_W(\bar{Y}_n, \hat{Y}_n) = \mathbf{E}(Y_n) \left| \frac{1}{\mathbf{E}(Y_n)} - \frac{2}{n\sigma^2} \right|.$$

Il faut ainsi estimer le terme d'erreur dans l'approximation de Kolmogorov

$$\frac{1}{\mathbf{E}(Y_n)} = \mathbf{P}(Z_n > 0) \sim \frac{2}{n\sigma^2}.$$

A partir d'une référence, les auteurs assurent que le terme d'erreur est en  $O\left(\frac{\log^2(n)}{n}\right)$ . Sur cette base, le deuxième terme de 37 domine le premier est conduit à une erreur en  $O\left(\frac{\log^2(n)}{n}\right)$ . Je pense qu'en fait l'erreur est en  $O\left(\frac{\log(n)}{n}\right)$ . Sur cette base, le deuxième terme est du même ordre de grandeur que le premier est conduit à une erreur en  $O\left(\frac{\log(n)}{n}\right)$ , comme pour le théorème précédent. Des investigations numériques vont dans notre sens (voir figure 6). Nous reportons en annexe le raisonnement sujet à validation.

### Annexe - Reste dans l'estimation de Kolmogorov

The proof of the Kolmogorov estimate (theorem 7) results on basic manipulations on recursive sequences. Indeed, the fundamental property of Galton-Watson process is the expression of the generating function  $f_n(s) := \mathbf{E}(s^{Z_n})$  in term of the generating function of the reproduction law

$$f_n(s) = f^{\circ n}(s) \quad (38)$$

Thus,  $u_n := \mathbf{P}(Z_n > 0)$  is a recursive sequence with  $u_0 = 1$  and  $u_{n+1} = \varphi(u_n)$  where  $\varphi(u) = 1 - f(1 - u)$ . We known that  $u_n \rightarrow 0$  i.e. a critical Galton-Watson process extinct almost surely. To find an equivalent of  $u_n$ , the classical approach consist in writing a Taylor expansion of  $\varphi$  at 0, and then to find a good transformation  $T$  such that the difference  $T(v_{n+1}) - T(v_n)$  is made of well-controllably terms. In particular, such as the difference has a limit when  $n \rightarrow \infty$ . Thereby, summing the equality of  $n$ , the left-hand term will telescope whereas the right-hand term can be analysed using basic facts about asymptotic behaviour of divergent series. In our case, we start to write the Taylor Expansion of  $f$  in 1. Because,  $f$  has a finite second moment we get

$$f(s) = 1 - (1 - s) + \frac{\sigma^2}{2}(1 - s)^2 + (1 - s)^2 \alpha(1 - s) \quad (39)$$

where  $\alpha$  is a continuous function on  $[0, 1]$  with  $\alpha(0) = 0$ . Thus we get the Taylor expansion of  $\varphi$  at 0

$$\varphi(u) = u - \frac{\sigma^2}{2}u^2 + u^2 \alpha(u) \quad (40)$$

Now, in view of the equivalent we want to find, we naturally consider the transformation  $T : u \mapsto u^{-1}$ . Applying  $T$  in the previous expression we obtain

$$\frac{1}{\varphi(u)} - \frac{1}{u} = \frac{\sigma^2}{2} + \beta(u) \quad (41)$$

where  $\beta$ , as  $\alpha$ , is a continuous function on  $[0, 1]$  with  $\beta(0) = 0$ . Injecting  $u_n$  in the previous expression we obtain

$$\frac{1}{u_{n+1}} - \frac{1}{u_n} = \frac{\sigma^2}{2} + \beta(u_n). \quad (42)$$

The right-hand side converge to  $\sigma^2/2$  as  $n \rightarrow \infty$ . Applying Cesaro Lemma we obtain

$$\frac{1}{n} \left( \frac{1}{u_n} - 1 \right) \rightarrow \frac{\sigma^2}{2} \quad (43)$$

from which follow the Kolmogorov estimate.

To find the following term in the expansion (??) we push the Taylor development to the next term. We have to suppose that the reproduction law has a third moment.

$$1 - f(s) = (1 - s) - \frac{\sigma^2}{2}(1 - s)^2 + \frac{\rho}{6}(1 - s)^3 + (1 - s)^3 \alpha(1 - s) \quad (44)$$

where  $\rho := f^{(3)}(1)$  is the third factorial moment of  $f$  and  $\alpha$  is a continuous function on  $[0, 1]$  with  $\alpha(0) = 0$ . Thus we get the Taylor expansion of  $\varphi$  at 0

$$\varphi(u) = u - \frac{\sigma^2}{2}u^2 + \frac{\rho}{6}u^3 + u^3\alpha(u). \quad (45)$$

Applying the same transform as before we get

$$\frac{1}{\varphi(u)} - \frac{1}{u} = \frac{\sigma^2}{2} + \left(\frac{\sigma^4}{4} - \frac{\rho}{6}\right)u + u\beta(u) \quad (46)$$

where  $\beta$ , so as  $\alpha$ , is a continuous function with  $\beta(0) = 0$ . Injecting  $u_n$  we obtain

$$\frac{1}{u_{n+1}} - \frac{1}{u_n} = \frac{\sigma^2}{2} + \left(\frac{\sigma^4}{4} - \frac{\rho}{6}\right)u_n + u_n\beta(u_n). \quad (47)$$

and summing over  $n$ ,

$$\frac{1}{u_n} = 1 + n\frac{\sigma^2}{2} + \left(\frac{\sigma^4}{4} - \frac{\rho}{6}\right)\sum_{k=0}^{n-1}u_k + \sum_{k=0}^{n-1}u_k\beta(u_k). \quad (48)$$

First, since  $u_n \sim \frac{2}{n\sigma^2}$  when  $n \rightarrow \infty$  we have

$$\sum_{k=0}^{n-1}u_k \underset{n \rightarrow \infty}{\sim} \frac{2}{\sigma^2} \log(n). \quad (49)$$

Then since  $u_n\beta(u_n) = o\left(\frac{1}{n}\right)$  we have

$$\sum_{k=0}^{n-1}u_k\beta(u_k) \underset{n \rightarrow \infty}{=} o(\log(n)). \quad (50)$$

Thereby,

$$\frac{1}{u_n} = n\frac{\sigma^2}{2} + \frac{2}{\sigma^2} \left(\frac{\sigma^4}{4} - \frac{\rho}{6}\right) \log(n) + o(\log(n)). \quad (51)$$

Finally, by inverting the expansion we get (52).

**Théorème 20.** *Let  $Z_n$  be a critical Galton-Watson process of law  $f$ . Suppose that  $f$  has a third moment. Then,*

$$\mathbf{P}(Z_n > 0) = \frac{2}{n\sigma^2} + \frac{2}{\sigma^2} \left(\frac{2\rho}{3\sigma^4} - 1\right) \frac{\log(n)}{n^2} + o\left(\frac{\log(n)}{n^2}\right) \quad (52)$$

where  $\rho := f^{(3)}(1)$  is the third factorial moment of  $f$ . In particular,

$$\mathbf{P}(Z_n > 0) = \frac{2}{n\sigma^2} + O\left(\frac{\log(n)}{n^2}\right). \quad (53)$$

In this case, the second bound in theorem 1.2 will be improved and thus unify the second bound with the first one.

**Théorème 21.** *Under the same hypothesis there is a constant  $C > 0$  such that*

$$d_W\left(\frac{2Z_n}{n\sigma^2}, \text{Exp}\right) \leq C \frac{\log(n)}{n}. \quad (54)$$

I did some investigations to check the formula above. For a linear fractional Galton-Watson process, we can have that  $2\rho = 3\sigma^4$  so that the second term is cancelled. We turn out to numerical simulations. The figure 6 shows the result we obtain for the binary reproduction law  $f(s) = \frac{1}{2} + \frac{1}{2}s^2$ . It appears that the error  $R_n = n \times \{n\mathbf{P}(Z_n > 0) - \sigma^2\}$  is logarithmic in  $n$  with a coefficient approximately equal to  $-2 = -\sigma^2$  (in this case  $\rho = 0$ ).

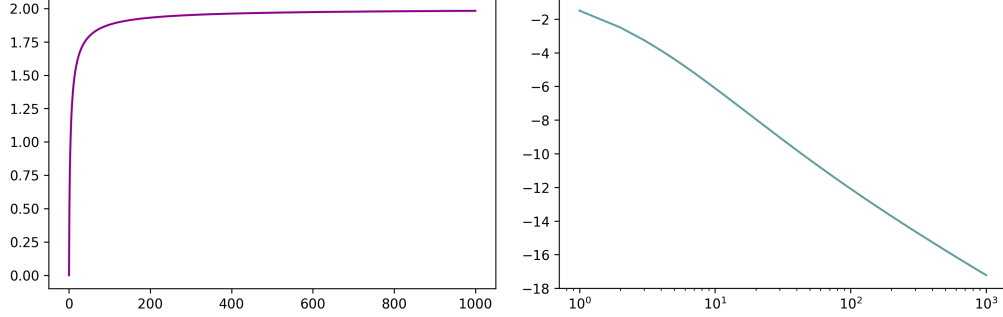


FIGURE 6 – The left-hand graphic shows the convergence of  $n\mathbf{P}(Z_n > 0)$  to  $\sigma^2 = 2$  et the right-hand graphic represent the error  $n \times \{n\mathbf{P}(Z_n > 0) - \sigma^2\}$  display in log-xscale. The calculations of  $\mathbf{P}(Z_n > 0)$  have been realized using the formula  $\mathbf{P}(Z_n > 0) = f^{\circ n}(0)$ .

## 4 Processus de Galton-Watson à environnement variable

Le processus de Galton-Watson a été très étudié et a connu diverses généralisations. Une première perspective de généralisation est de considérer différents "types" d'enfants. Par exemple, en attribuant un sexe aux individus. On parle généralement de processus de Galton-Watson *multi-type*. Mais surtout le point susceptible de généralisation est la loi de reproduction. On peut ajouter une dépendance en la génération de loi de reproduction c'est à dire choisir pour chaque génération une loi de reproduction propre. L'ensemble de ces lois de reproductions forme l'*environnement du processus* et on parle généralement de processus de Galton-Watson à environnement variable. On peut aussi ajouter de l'aléa dans le choix de la loi de reproduction. La loi de reproduction est alors variable aléatoire à valeurs dans l'ensemble des distributions de probabilités sur  $\mathbf{N}$ . On parle alors de *processus de Galton-Watson en environnement aléatoire*. Ces généralisations permettent alors une description plus fidèle des phénomènes que l'on cherche à modéliser. Pour notre part, nous nous concentrerons sur la généralisation à un environnement aléatoire. Dans la première partie, nous définissons de telles processus et en donnons les propriétés élémentaires notamment l'expression de la fonction génératrice qui se généralise naturellement. Ainsi, l'étude se ramène à celle de la composition successive des fonctions génératrices. Cette étude semble avoir eu du mal à se lancer à cause de certains exemples pathologiques. Mais on découvre que sous certaines hypothèses plus ou moins restrictives, ces cas pouvaient être écartés et un comportement général ressortait. Nous présentons l'approche de G.KERSTING dans [Ker20]. C'est un travail récent qui généralise une part des approches précédentes et sera bien adaptée dans le cadre de la prochaine section. Son point de vue repose sur la notion de *fonction de forme*, présentée dans la deuxième partie, qui s'appuie sur des idées déjà présentes dans le cas classique. Dans la troisième partie, nous montrerons la convergence du processus dû à LINDVALL et CHURCH. Nous verrons alors apparaître un nouveau régime où le processus se stabilise sur une valeur finie non nulle. Dans la quatrième partie, nous étudierons à quelle condition le processus s'éteint. Dans la cinquième, nous tenterons de généraliser les résultats de YAGLOM pour enfin dans la dernière étudier la vitesse de convergence vers la limite de Yaglom, objet de ce stage.

### 4.1 Définition du processus

Un environnement est une suite  $f = (f_n)_{n \geq 1}$  de mesures de probabilités sur  $\mathbf{N}_0$  où  $f_n$  représente la loi de reproduction des individus de la génération  $n$ . On généralise facilement la définition du processus de Galton-Watson classique (1) au cas où les lois de reproductions changent, il s'agit juste d'ajouter à la loi des  $X_{i,n}$  la dépendance en la génération. La forme du processus reste donc la même, et l'on a ainsi toujours une expression exacte de la fonction génératrice du processus : c'est la composée des fonctions génératrices des lois de reproductions des générations antérieures (proposition 22). Cette expression permet de dériver les moments du processus (corollaire 23). Finalement, le but de cette partie est juste de fixer les notations.

**Définition 7.** *Le processus de Galton-Watson en l'environnement variable  $f$  de population initiale  $Z_0$  est la suite  $(Z_n : n \geq 0)$  définie par son premier terme  $Z_0$ , variable aléatoire sur  $\mathbf{N}_0$ , et la relation de récurrence*

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n}, \quad (55)$$

où les  $X_{i,n} - (i,n) \in \mathbf{N}_0 \times \mathbf{N}_0$  - sont des variables aléatoires indépendantes de lois  $X_{i,n} \sim f_n$ . Lorsque  $Z_0 = 1$  p.s nous qualifierons le processus de standard.

**Proposition 22.** *La fonction génératrice  $f^n$  de  $Z_n$  est*

$$f^n(s) = f_1 \circ \dots \circ f_n(s). \quad (56)$$

Plus généralement, la fonction génératrice de  $Z_n | Z_k = 1$  est,

$$f_{k:n}(s) := f_{k+1} \circ \dots \circ f_n(s). \quad (57)$$



Dans toute la suite nous ferons l'hypothèse suivante.

**Hypothèse 0.** Les lois de reproductions  $f_n$  ont toutes une moyenne non nulle et une variance finie.<sup>7</sup> Nous noterons alors la moyenne, le deuxième moment factoriel normalisé de  $f_k$  et la variance normalisée de  $f_k$  de la façon suivante

$$\begin{aligned}\mu_k &:= f'_k(1) = \mathbf{E}[X_k] \\ \nu_k &:= \frac{f''_k(1)}{f'_k(1)^2} = \frac{\mathbf{E}[X_k(X_k - 1)]}{\mathbf{E}[X_k]^2} \\ \sigma_k &:= \nu_k + \frac{1}{\mu_k} - 1 = \frac{\mathbf{Var}(X_k)}{\mathbf{E}[X_k]^2}.\end{aligned}$$

**Corollaire 23.** *Moments du processus*

$$\mu_{k:n} := f'_{k:n}(1) = \prod_{l=k+1}^n \mu_l \quad (58)$$

$$\nu_{k:n} := \frac{f''_{k:n}(1)}{f'_{k:n}(1)^2} = \sum_{l=k+1}^n \frac{\nu_l}{\mu_{k:l-1}}. \quad (59)$$

$$\sigma_{k:n} := -1 + \frac{f''_{k:n}(1) + f'_{k:n}(1)}{f'_{k:n}(1)^2} = \sum_{l=k+1}^n \frac{\sigma_l}{\mu_{k:l-1}}. \quad (60)$$

sont respectivement le deuxième moment factoriel normalisé de  $f_k$  et la variance normalisée de  $f_{k:n}$ . En particulier, voici les premiers moments de notre processus

$$\mu^n := \mu_1 \cdots \mu_n, \quad (61)$$

$$\nu^n := \sum_{k=1}^n \frac{\nu_k}{\mu^{k-1}}, \quad (62)$$

$$\sigma^n := \sum_{k=1}^n \frac{\sigma_k}{\mu^{k-1}}. \quad (63)$$

Pour récapituler nous employons un indice bas  $k$  pour désigner les lois de reproduction, par un double indice bas  $k : n$  pour désigner l'itéré des lois de reproductions à partir de  $k + 1$  jusqu'à  $n$  et un indice haut pour désigner le processus que l'on pourrait aussi noter  $0 : n$ .

**Remarque.** *Arbre généalogique*

Comme dans le cas univarié, un processus de Galton-Watson standard à environnement variable peut se voir comme une distribution sur l'espace des arbres  $\mathcal{T}$ . En posant,

$$\text{GW}_f^n(t) = \prod_{u \in t : |u| \leq n} f_{|u|}(k_u(t)) \quad (64)$$

on définit comme dans (\*) une suite de mesure sur la filtration  $\mathcal{F}_n$  vérifiant la condition de compatibilité. Ainsi, le théorème d'extension de Kolmogorov donne l'existence d'une mesure  $\text{GW}_f$  sur  $\mathcal{T}$  dont les marginales vérifient (64). La mesure obtenue dépend de l'environnement  $f$ . Alors, sous  $\text{GW}_f$  les variables aléatoires  $Z_n : t \mapsto Z_n(t)$  forment un processus de Galton-Watson standard d'environnement  $f$ .

## 4.2 Fonctions de formes

main l'expression  $f_1 \circ f_2 \circ \cdots \circ f_n(s)$ . Dans le cas constant, cette expression se simplifie considérablement : il n'y a plus qu'une fonction dont il s'agit d'étudier la dynamique. Dès lors, beaucoup de résultats, comme ceux concernant la probabilité d'extinction, tombaient relativement simplement

7. Le fait de demander la moyenne non nulle revient à exclure la loi de reproduction  $\delta_0$ .

par une étude élémentaire de la fonction  $f$ . Cette démarche n'est plus possible, il faut utiliser des techniques plus avancées. Dans le cas classique, lorsque regarder simplement  $f^{\circ n}$  ne suffisait pas à obtenir le résultat, il apparût judicieux de considérer l'expression

$$\frac{1}{1-f(s)} - \frac{1}{\mu(1-s)}.$$

C'est l'idée symbolique à la base des résultats de Kolmogorov (théorème 7) et de Yaglom (théorème 8). Ici, nous rajoutons la moyenne pour inclure tous les cas, pas seulement ceux où la moyenne vaut 1. Nous noterons généralement  $\varphi(s)$  la fonction précédente. Ainsi,

$$\frac{1}{1-f(s)} = \frac{1}{\mu(1-s)} + \varphi(s).$$

Cette décomposition correspond en quelque sorte à une décomposition en un terme de moyenne et le terme complémentaire qui correspond ainsi à la "forme" :  $\varphi$  est la *fonction de forme* associée à  $f$ . Selon [Ker20], nous partirons de cette idée pour notre étude. Nous emploierons les mêmes conventions de notations que les fonctions génératrices. Ainsi,  $\varphi_k$  désigne la fonction de forme associée à  $f_k$ ,  $\varphi_{k:n}$  est la fonction de forme associée à  $f_{k:n}$  et  $\varphi^n$  est la fonction de forme associée à  $f^n$ . De la proposition 22 nous déduisons l'expression suivant des fonctions de formes.

**Corollaire 24.** *Fonctions de formes*

$$\varphi_{k:n}(s) = \sum_{l=k+1}^n \frac{\varphi_l(f_{l:n}(s))}{\mu^{l-1}}. \quad (65)$$

Une fonction de forme est bien définie et continue sur  $[0, 1)$ , mais elle se prolonge par continuité en 1 et fait apparaître le deuxième moment factoriel  $\mu$  de la loi :

$$\lim_{s \uparrow 1} \varphi(s) = \frac{\nu}{2}.$$

En particulier, en évaluant en 1 l'équation (65) on retrouve la formule (59). Notons de plus que par convexité des fonctions génératrices, une fonction de forme est toujours positive.

### 4.3 Convergence

On commence par étudier la convergence de  $Z_n$ . Dans le cas variable, les valeurs possibles de  $Z_n$  à l'infini est plus riche ; on a plus la dichotomie extinction/explosion. Il peut arriver que  $Z_n$  se stabilise en un état positif. Par contre, le processus ne peut osciller entre différents états positifs, le processus convergence presque sûrement c'est à dire que étant à valeurs entières, il atteint au bout d'un certain temps un état sur lequel il se stabilise ou diverge vers l'infini. Ce résultat est remarquable car il ne demande aucune hypothèse en plus sur notre processus.

**Théorème 25.** *Il existe une variable aléatoire  $Z_\infty$  à valeurs dans  $\bar{\mathbf{N}}$  tel que*

$$Z_n \longrightarrow Z_\infty \quad p.s.$$

Il est également possible de caractériser le cas où  $Z_\infty$  prend d'autres valeurs que 0 ou  $\infty$ . Cela correspond à une processus où les enfants tendent à être *enfants uniques* en un certain sens.

**Théorème 26.** *La variable aléatoire limite  $Z_\infty$  prend une valeur distincte de 0 et  $\infty$  avec une probabilité non nulle ssi  $\sum_{n=1}^{\infty} (1 - f_n[1]) < \infty$ . Dans ce cas, on dit que le processus est asymptotiquement dégénéré.*

**Exemple. Branchement symétrique**

Considérons l'environnement aléatoire  $f = (f_n)_{n \geq 1}$  avec,

$$f_n[0] = f_n[2] = \frac{1}{2n^\alpha}, \quad f_n[1] = 1 - \frac{1}{n^\alpha}.$$

Alors,

$$\sum_{n=1}^{\infty} (1 - f_n[1]) = \sum_{n=1}^{\infty} \frac{1}{n^\alpha} = \begin{cases} < \infty & \text{si } \alpha \leq 1 \\ +\infty & \text{si } \alpha > 1 \end{cases}.$$

Ainsi, si  $\alpha > 1$  le processus est asymptotiquement dégénérée.

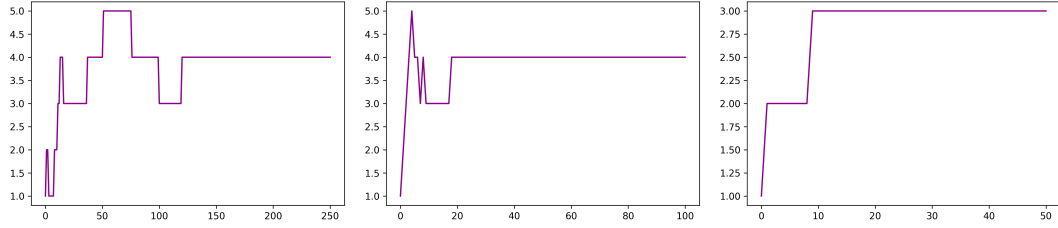


FIGURE 7 – Simulations processus de Galton-Watson à environnement variable pour le branchement symétrique présentée dans l'exemple précédent avec de gauche à droite  $\alpha = 1.2, 1.5$  et  $2$ .

#### 4.4 Extinction

On s'intéresse maintenant au phénomène d'*extinction*. Dans la sous-section précédente, nous avons établie la convergence presque sûre de notre processus. On cherche à comprendre quand est ce que cette limite est presque sûrement nulle dans l'optique de déterminer les régimes du processus, généralisant ce que l'on sait du processus de Galton-Watson classique.

On suit la démarche de [Ker20]. L'idée fondamentale est la suivante : pour qu'il y ait extinction du processus, il faut que l'écart relatif de la génération  $n$  à sa moyenne diverge lorsque  $n \rightarrow \infty$  c'est à dire ;

$$\frac{\sqrt{\text{Var}(Z_n)}}{\mathbf{E}[Z_n]} \xrightarrow{n \rightarrow \infty} \infty.$$

Intuitivement, cela signifie que le processus oscille fortement par rapport à sa moyenne ; il peut très bien être nettement au dessus de sa moyenne mais aussi nettement en dessous, or si le processus tombe sur 0, c'est irréversible, la population s'est éteinte. Ainsi, [Ker20] établit que

$$\mathbf{P}(Z_n > 0) \asymp \frac{(\mathbf{E}Z_n)^2}{\mathbf{E}(Z_n^2)}.$$

D'un côté, la borne inférieure découle de manipulations standards, c'est un fait général pour les lois à valeurs entières. De l'autre côté, pour la majoration, nous avons besoin d'hypothèses supplémentaires. Suivant la démarche présentée, on passe par les fonctions de formes ; on réécrit

$$\mathbf{P}(Z_n > 0)^{-1} = (1 - f_n(0))^{-1} = \frac{1}{\mu^n} + \varphi^n(0).$$

Sachant que  $\frac{\mathbf{E}[Z_n(Z_n-1)]}{\mathbf{E}[Z_n]^2} = \varphi^n(1)$ , il nous faut une minoration de  $\varphi^n(0)$  par rapport à  $\varphi^n(1)$  uniformément en  $n$ . En regardant la forme de  $\varphi^n$ , on voit qu'il suffit d'avoir une minoration de  $\varphi^n(s)$  par rapport à  $\varphi^n(1)$  uniforme en  $n$  et  $s$ . En fait, les fonctions de formes vérifient de façon générale l'inégalité

$$\frac{1}{2}\varphi(0) \leq \varphi(s) \leq 2\varphi(1), \quad (66)$$

pour tout  $s \in [0, 1]$ , de sorte qu'il suffit l'inégalité pour  $s = 0$ .

**Hypothèse A.** On suppose que l'environnement aléatoire  $\mathbf{f}$  est telle que  $\varphi_n(0) \geq c\varphi_n(1)$  pour une constante  $c > 0$  indépendante de  $n$ . En écrivant les expressions des fonctions de forme en fonctions des distributions associés, cette inégalité s'écrit en terme de lois de reproductions :

$$\mathbf{E}(X_n(X_n - 1)) \leq c\mathbf{E}(X_n | X_n \geq 1)\mathbf{E}(X_n \mathbf{1}_{X_n \geq 2}).$$

Notons que comme  $\frac{1}{2}\mathbf{E}(X_n^2 \mathbf{1}_{X_n \geq 2}) \leq \mathbf{E}(X_n(X_n - 1)) \leq \mathbf{E}(X_n^2 \mathbf{1}_{X_n \geq 2})$  le terme de droite peut encore être remplacé par  $\mathbf{E}(X_n^2 \mathbf{1}_{X_n \geq 2})$ .

Dans ce cas, nous aurons

$$\mathbf{P}(Z_n > 0)^{-1} \geq \min(1, c) \left\{ \frac{1}{\mu^n} + \frac{\mathbf{E}[Z_n(Z_n - 1)]}{\mathbf{E}[Z_n]^2} \right\} = \min(1, c) \frac{\mathbf{E}[Z_n^2]}{\mathbf{E}[Z_n]^2}.$$

**Théorème 27.** *Sous l'hypothèse A, le processus s'éteint presque sûrement ssi  $\mathbf{E}(Z_n) = o(\sqrt{\mathbf{Var}(Z_n)})$  c'est à dire*

$$\sum_{k=1}^{\infty} \frac{\rho_k}{\mu_{k-1}} = \infty.$$

**Remarque.** L'hypothèse A est satisfaite pour des distributions usuelles : Poisson, binomiale, géométrique, hypergéométrique, linéaire fractionnaire ou encore binomiale négative sans aucune restriction sur les paramètres. Plus précisément, pour  $X$  une variable aléatoire de l'une de ces familles, on a l'inégalité

$$\mathbf{E}[X^2 \mathbf{1}_{X \geq 2}] \leq c\mathbf{E}[X | X \geq 1]\mathbf{E}[X \mathbf{1}_{X \geq 2}]$$

pour une constante  $c$  ne dépendant que de la famille. Ainsi, si l'on choisit un environnement composés de ces lois usuelles, l'hypothèse A sera toujours vérifiée. L'inégalité précédente est aussi vérifié dès lors que  $X$  est à support borné ; alors la constante  $c$  peut être choisie de façon à ne dépendre que du support de  $X$ . On peut ainsi ajouter à notre environnement un ensemble de loi à support uniformément borné. Toutefois, tous les environnements ne satisfont par l'hypothèse A. Par exemple

$$f_n \propto \frac{1}{k^\alpha} \mathbf{1}_{k \geq 2}$$

pour  $\alpha \in (2, 3]$ .

**Remarque.** *Croissance du processus*

L'hypothèse A contraint la croissance du processus sur l'évènement de non extinction ;  $\mu$  est le bon taux de croissance (voir [Ker20] théorème 2)

$$W_n = \frac{Z_n}{\mu_n} \xrightarrow{\text{p.s.}} W \quad \text{avec} \quad \mathbf{P}(W = 0) = q.$$

Il est cependant à noter que MACPHEE et SCHUH ont construit un processus de Galton-Watson à environnement variable avec différents taux de croissances [MS08].

## 4.5 Limite de Yaglom

On termine par étudier le comportement vers l'extinction en étudiant le processus  $Y_n = Z_n | Z_n > 0$  en cherchant à généraliser les résultats du cas classique. Dans le cas sous-critique, il n'est pas espérable d'avoir une unique loi limite en général. En effet, si l'on choisit  $f_1$  et  $f_2$  deux distributions sous-critiques puis qu'on considère l'environnement aléatoire

$$f_n = \begin{cases} f_1 & \text{si } n_{2k} < n \leq n_{2k+1} \\ f_2 & \text{sinon} \end{cases}$$

où  $(n_k)_{k \geq 1}$  est une suite strictement croissante d'entiers naturelles. Si les écarts  $n_{k+1} - n_k$  tendent vers l'infini, on voit que  $Z_{n_{2k+1}} | Z_{n_{2k+1}} > 0$  convergera vers la limite de Yaglom associé au processus

de Galton-Watson classique de loi de reproduction  $f_1$  tandis que  $Z_{n_{2k}} | Z_{n_{2k}} > 0$  convergera vers celle de  $f_2$ . Dès lors que  $f_1$  et  $f_2$  ont une limite de Yaglom différente, on aura différentes lois limites. Le problème précédent disparaît dans le cas critique, car alors la limite de Yaglom est universelle. On peut alors espérer obtenir une loi limite, et c'est justement ce que nous allons voir dans la suite. Pour se faire, on suit la démarche de [Ker20] qui généralise la preuve usuellement présentée dans le cas classique (voir annexe). Pour faire cela il lui faut une hypothèse plus forte (B). Sous cette hypothèse, KERSTING établit la convergence du processus convenablement normalisée vers la loi exponentielle standard.

**Hypothèse B.** L'environnement  $\mathbf{f}$  est tel que pour tout  $\varepsilon > 0$ , il existe une constante  $c_\varepsilon$  indépendante de  $n$  tel que

$$\mathbf{E}(X_n^2; X_n \geq c_\varepsilon(1 + \mathbf{E}X_n)) \leq \varepsilon \mathbf{E}(X_n^2; X_n \geq 2). \quad (67)$$

**Théorème 28.** En supposant que  $b_n = \mathbf{E}[Z_n | Z_n > 0] \rightarrow \infty^8$  on a,

$$\frac{1}{b_n} Z_n | Z_n > 0 \rightarrow \text{Exp}.$$

De plus,

$$b_n \sim \frac{\mu^n}{2} \nu^n = \frac{\mu^n}{n} \sum_{k=1}^n \frac{\nu_k}{\mu^{k-1}} \quad (68)$$

et l'on peut remplacer  $b_n$  par son équivalent dans la limite précédente.

L'hypothèse B est une généralisation de l'hypothèse A. Sous l'hypothèse B nous avons

$$\mathbf{E}(X_n^2; 2 \leq X_n) \leq 2c_{1/2}(1 + \mathbf{E}X_n)\mathbf{E}(X_n; 2 \leq X_n) \quad (69)$$

Et comme  $1 + \mathbf{E}X_n \leq 2\mathbf{E}(X_n | X_n \geq 1)$ , nous avons bien (A) avec  $c_A = 4c_{1/2}$ . En effet,

$$\begin{aligned} \mathbf{E}(X_n^2; 2 \leq X_n) &\leq \mathbf{E}(X_n^2; 2 \leq X_n \leq c_{1/2}(1 + \mathbf{E}X_n)) + \mathbf{E}(X_n^2; X_n > c_{1/2}(1 + \mathbf{E}X_n)) \\ &\leq \mathbf{E}(X_n^2; 2 \leq X_n \leq c_{1/2}(1 + \mathbf{E}X_n)) + \frac{1}{2}\mathbf{E}(X_n^2; 2 \leq X_n) \end{aligned}$$

d'où

$$\mathbf{E}(X_n^2; 2 \leq X_n) \leq 2\mathbf{E}(X_n^2; 2 \leq X_n \leq c_{1/2}(1 + \mathbf{E}X_n)). \quad (70)$$

A l'instar de A l'hypothèse B est satisfaite par la plupart des familles de probabilités usuelles : Poisson, binomiale, hypergéométrique, binomiale négative et à support dans  $\{0, \dots, N\}$  à l'exception des lois linéaires fractionnaires. En fait, il est audacieux de vouloir vérifier B telle quelle. KERSTING énonce ainsi une autre hypothèse, se basant sur l'existence d'un moment d'ordre 3, qui implique B. Avec cette hypothèse C on peut simplement vérifier que les lois usuelles satisfont à nos hypothèses. Pour plus de détails voir [KV17].

**Hypothèse C.** Il existe une constante  $c$  tel que

$$f_n'''(1) \leq c f_n''(1)(1 + f_n'(1)).$$

## 4.6 Vitesse de convergence

En suivant l'article [CTJP23], nous allons généralisé les bornes obtenues sur la convergence vers la limite de Yaglom au cas d'un environnement variable. Naturellement, on a vu des résultats de la partie précédente et de la nécessité d'avoir un moment d'ordre 3, nous allons supposer que notre environnement satisfait à l'hypothèse C. En suivant stricto sensu la même démarche que dans la partie 3.2, on montre le résultats suivant.

---

8. Cela correspond au régime critique en environnement variable

**Théorème 29.** Soit  $Z_n = Z_n^{\mathbf{f}}$  un processus de Galton-Watson critique en environnement variable  $\mathbf{f}$  satisfaisant C. Il existe une constante  $C$  tel que

$$d_W\left(\frac{1}{b_n}Z_n \mid Z_n > 0, \text{Exp}\right) \leq C \left( \frac{1}{\mu^n \nu^n} + \frac{1}{\nu^n} \sum_{j=1}^{n-1} \frac{\nu_j}{\mu^{j-1}} \frac{1+\mu_j}{\mu^j(\rho^n - \rho^j)} + \frac{\nu_n}{\mu^{n-1}}(1+\mu_n) \right). \quad (71)$$

Telle quelle, la majoration 71 est difficilement exploitable. Dans l'article [CTJP23], les auteurs travaillent cette borne afin de la simplifier.

**Corollaire 30.** Supposons que l'environnement  $\mathbf{f}$  est tel  $\inf_{n \in \mathbf{N}^*} f_n''(1) > 0$ . Alors il existe une constante  $C > 0$  telle que

$$d_W\left(\frac{1}{b_n}Z_n \mid Z_n > 0, \text{Exp}\right) \leq C(1 + M_n)^5 \left( \frac{\log(\mu^n \nu^n) + \log(\mu_n)}{\mu^n \nu^n} + \frac{s_n}{\nu^n} \right) \quad (72)$$

où  $M_n = \sup_{1 \leq k \leq n} \mu_k$  et,

$$s_n = \sum_{k=2}^{n-1} (\log(\nu^k \mu^k) + \log(\mu_k)) \left| \frac{1}{\mu^{k-2}} - \frac{1}{\mu^k} \right|. \quad (73)$$

De plus, si  $\sup_{n \in \mathbf{N}^*} f_n''(1) < \infty$  alors,

$$d_W\left(\frac{1}{b_n}Z_n \mid Z_n > 0, \text{Exp}\right) \leq C(1 + M_n)^5 \left( \frac{\log(\mu^n \nu^n)}{\mu^n \nu^n} + \frac{s_n}{\nu^n} \right). \quad (74)$$

Sous cette forme, on voit apparaître la forme de la borne qu'on a obtenu dans le cas classique. D'ailleurs, si l'on cherche à appliquer ce résultat à un environnement constant, on retrouve bien la borne en  $\frac{\log(n)}{n}$ .

## Conclusion

Ce stage m'aura permis d'approfondir mes connaissances sur le processus de Galton-Watson et d'entrevoir la recherche sur le sujet. J'ai ainsi pu mieux comprendre les motivations de la recherche mathématique et comment cette dernière s'exerce. D'un autre côté, ce sujet classique m'a fait questionner la présentation qu'on en fait, notamment celle que j'ai reçu en cours. J'ai ainsi cherché à décrypter le sujet, désosser les preuves pour saisir tous les rouages du mécanisme. Cela m'aura permis d'éclaircir certains faits alors obscurs jusqu'alors. Cette démarche semble avoir porté ses fruits ; c'est l'analyse de l'estimation de Kolmogorov qui m'a conduit au petit travail personnel présenté en fin de section 3. En attendant la suite.

## Annexes - Démonstrations des théorèmes

### Théorème de convergence de Yaglom critique - Théorème 8

On calcule la transformée de Laplace<sup>9</sup> de  $\tilde{Z}_n$

$$\mathbf{E} \left[ e^{-u\tilde{Z}_n} \right] = \frac{\mathbf{E} \left[ e^{-uZ_n/n} \mathbf{1}_{Z_n > 0} \right]}{\mathbf{P}(Z_n > 0)} = \frac{f_n(e^{-u/n}) - f_n(0)}{1 - f_n(0)} = 1 - \frac{1 - f_n(e^{-u/n})}{1 - f_n(0)}.$$

Il faut ainsi estimer l'écart de  $f_n(s)$  à 1. Pour le dénominateur,  $s = 0$  et nous sommes ramenés à la partie précédente. On sait alors que

$$1 - f_n(0) \sim \frac{2}{n\sigma^2}.$$

Pour le numérateur  $s = e^{-u/n}$ . On va réutiliser les idées précédentes pour s'en sortir. On peut écrire avec les notations précédentes

$$\frac{1}{n(1 - f_n(e^{-u/n}))} = \frac{1}{n(1 - e^{-u/n})} + \frac{\sigma^2}{2} + \frac{1}{n} \sum_{k=0}^{n-1} \beta(f^k(e^{-u/n})).$$

Le premier terme tend vers  $\frac{1}{u}$  lorsque  $n \rightarrow \infty$ . Pour le second terme, nous allons voir qu'il tend encore une fois vers 0. Dorénavant, a une dépendance en  $n$  dans la somme mais aussi dans la variable. Une façon de générer ces doubles limites est d'avoir la convergence uniforme de la suite de fonctions. Et, comme  $f$  est croissante,  $f^k(0) \leq f^k(s) \leq 1$  donc  $f^k(s)$  converge uniformément vers 1 sur  $[0, 1]$  puis comme  $\beta$  est continue,  $\beta(f^k(s))$  converge uniformément vers  $\beta(1) = 0$  sur  $[0, 1]$ . Ainsi, par le lemme de Cesàro  $\frac{1}{n} \sum_{k=0}^{n-1} \beta(f^k(s))$  converge uniformément vers 0 sur  $[0, 1]$  d'où l'on déduit que

$$\frac{1}{n} \sum_{k=0}^{n-1} \beta(f^k(e^{-u/n})) \rightarrow 0.$$

Finalement,

$$\frac{1}{n(1 - f_n(e^{-u/n}))} \rightarrow \frac{1}{u} + \frac{\sigma^2}{2}$$

puis,

$$\begin{aligned} \mathbf{E} \left[ e^{-u\tilde{Z}_n} \right] &\rightarrow 1 - \frac{\frac{\sigma^2}{2}}{\frac{1}{u} + \frac{\sigma^2}{2}} = \frac{1}{1 + \frac{u\sigma^2}{2}} \\ &= \frac{2}{\sigma^2} \int_0^\infty e^{-ux} e^{-x \frac{2}{\sigma^2}} dx. \end{aligned}$$

### Vitesse de convergence du cas classique - Théorème 19

En revenant à l'équation (21)

$$\mathbf{E}(|R_n - \tilde{R}_n|) \leq \sum_{j=1}^n \mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] + \mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}]. \quad (75)$$

Pour le premier terme on a selon (23) et (??)

$$\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \mathbf{E} \left[ I_j \mathbf{P}(A_{n,j}^c | I_j, N_j) \right]. \quad (76)$$

Maintenant nous allons majorer expliciter la probabilité, nous n'avions que affirmer que sa limite étant nulle jusqu'alors. On a sachant que  $Z_{n,j}$  est un GW de population initiale  $I_j$ ,

$$\mathbf{P}(A_{n,j}^c | I_j, N_j) = \mathbf{P}(L_{n,j} > 0 | I_j, N_j) \leq \mathbf{P}(Z_{n,j}^* | I_j, N_j) = I_j \mathbf{P}(Z_{n-j} > 0) \quad (77)$$

---

9. Le fait de considérer la transformée de Laplace plutôt que la fonction génératrice est artificielle. Cela simplifie juste les expressions.



ainsi,

$$\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \mathbf{E}[I_j^2] \mathbf{P}(Z_{n-j} > 0) \quad (78)$$

avec  $\mathbf{E}[I_j^2] = \mathbf{E}[X(X-1)^2] \leq \mathbf{E}[X^3]$ . Pour le second terme d'après 27,

$$\mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \sigma^2 \mathbf{P}(A_{n,j}^c) \quad (79)$$

donc avec ce qui précède

$$\mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \sigma^2 \mathbf{E}[I_j] \mathbf{P}(Z_{n-j} > 0) = \sigma^4 \mathbf{P}(Z_{n-j} > 0). \quad (80)$$

En combinant ces deux égalités on aboutit à

$$\mathbf{E}(|R_n - \tilde{R}_n|) \leq C_0 \sum_{j=1}^n \mathbf{P}(Z_{n-j} > 0) = C \sum_{k=0}^{n-1} \mathbf{P}(Z_k > 0) \quad (81)$$

où  $C$  est une constante. Par l'approximation de Kolmogorov (théorème 7),  $\mathbf{P}(Z_k > 0) \sim 2\sigma^{-2}k^{-1}$  donc,

$$\sum_{k=0}^{n-1} \mathbf{P}(Z_k > 0) \sim \frac{2}{\sigma^2} \log(n). \quad (82)$$

On a donc que  $\mathbf{E}(|R_n - \tilde{R}_n|) = O(\log(n))$ . Finalement, d'après 36 nous avons

$$d_W(\bar{Y}_n, \text{Exp}) = O\left(\frac{\log(n)}{n}\right).$$

## Preuve du critère d'extinction de Kersting - Théorème 27

Pour une variable aléatoire  $Z$  quelconque,

$$\mathbf{P}(Z > t\mathbf{E}Z) \geq \frac{\mathbf{E}(Z \mathbf{1}_{Z > t\mathbf{E}Z})^2}{\mathbf{E}(Z^2)} \quad (83)$$

en appliquant l'inégalité de Cauchy-Schwartz à  $Z \mathbf{1}_{Z > t\mathbf{E}Z}$ . Or,

$$\mathbf{E}(Z \mathbf{1}_{Z > t\mathbf{E}Z}) = \mathbf{E}(Z) - \mathbf{E}(Z \mathbf{1}_{Z \leq t\mathbf{E}Z}) \geq (1-t)\mathbf{E}(Z) \quad (84)$$

d'où

$$\mathbf{P}(Z > t\mathbf{E}Z) \geq (1-t)^2 \frac{\mathbf{E}(Z)^2}{\mathbf{E}(Z^2)}. \quad (85)$$

Cette inégalité est couramment appelée *inégalité de Paley-Zygmund*. Dans notre cas, en appliquant l'inégalité à  $Z_n$  et en faisant tendre  $t$  vers 0, on obtient

$$\mathbf{P}(Z_n > 0) \geq \frac{\mathbf{E}(Z_n)^2}{\mathbf{E}(Z_n^2)}. \quad (86)$$

Ce qui constitue bien la minoration recherchée. Pour la majoration, écrivons

$$\frac{1}{\mathbf{P}(Z_n > 0)} = \frac{1}{\mu_n} + \varphi_n(0). \quad (87)$$

Alors,

$$\varphi_n(0) = \sum_{k=1}^n \frac{\varphi_k(f_{k:n}(0))}{\mu^{k-1}} \geq \frac{1}{2} \sum_{k=1}^n \frac{\varphi_k(0)}{\mu^{k-1}} \geq \frac{c}{2} \varphi_n(1). \quad (88)$$

Ainsi,

$$\frac{1}{\mathbf{P}(Z_n > 0)} \geq \min(1, c/2) \left\{ \frac{1}{\mu^n} + \varphi_n(1) \right\}. \quad (89)$$

Comme  $\varphi_n(1) = \frac{\mathbf{E}(Z_n(Z_n-1))}{\mathbf{E}(Z_n)^2}$ , le terme entre crochet est  $\frac{\mathbf{E}(Z_n^2)}{\mathbf{E}(Z_n)^2}$ . Finalement nous avons montré qu'il existe une constante  $C$  tel que

$$\frac{\mathbf{E}(Z_n)^2}{\mathbf{E}(Z_n^2)} \leq \mathbf{P}(Z_n > 0) \leq C \frac{\mathbf{E}(Z_n)^2}{\mathbf{E}(Z_n^2)}. \quad (90)$$

On en déduit que les deux assertions suivantes sont équivalentes

- i.  $\mathbf{P}(Z_n > 0) \rightarrow 0$ ;
- ii.  $\mathbf{E}(Z_n)^2 = o(\mathbf{E}(Z_n^2))$ .

Mais, comme  $\mathbf{P}(Z_n > 0) = \mathbf{E}[\mathbf{1}_{Z_n > 0}] \rightarrow \mathbf{E}[\mathbf{1}_{Z_\infty > 0}] = \mathbf{P}(Z_\infty > 0)$  par convergence dominée, l'assertion i. correspond à ce que  $Z_\infty = 0$  p.s. Tandis que comme  $\mathbf{Var}(Z_n) = \mathbf{E}(Z_n^2) - \mathbf{E}(Z_n)^2$ ,

$$\frac{\mathbf{Var}(Z_n)}{\mathbf{E}(Z_n)^2} = \frac{\mathbf{E}(Z_n^2)}{\mathbf{E}(Z_n)^2} - 1$$

donc cette quantité tend vers l' $\infty$  si et ssi ii.

### Théorème de convergence de Lindvall et Church - Théorème 25

On suit la démonstration présentée dans [KV17]. On raisonne en supposant la non extinction du processus, c'est à dire en excluant le cas  $Z_n \rightarrow 0$  p.s. Une étape préliminaire est de montrer que la suite de fonctions génératrices  $f^n$  converge simplement sur  $[0, 1)$  vers une fonction  $f$  continue et strictement croissante. On reporte ce fait dans un lemme suivant la démonstration. L'idée est d'exploiter les s-martingales attachés au processus de la forme  $M_n = a_n^{Z_n}$ . Rappelons que si  $f_{n+1}(a_{n+1}) \geq a_n$ ,  $M_n$  est une sur-martingale (positive). En s'appuyant sur  $f$ , on peut construire une suite  $a_n$  de telle sorte que  $a_n \downarrow 1/2$  et  $f_{n+1}(a_{n+1}) \leq a_n$ . Pour cela, on peut considérer  $b_n$  une suite décroissante de réels avec  $f_n(1/2) \leq b_n \leq 1$  et  $b_n \downarrow f(1/2)$  puis définir  $a_n$  par l'équation suivante  $f^n(a_n) = b_n$ . Regardons la convergence de cette s-martingale. Comme  $M_n$  est une sur-martingale positive, elle converge presque sûrement vers une variable aléatoire  $M$ . Maintenant,  $a_n$  étant décroissante minorée, elle converge vers une limite  $a \in [0, 1)$ . Maintenant, par le *deuxième théorème de Dini*,  $f^n$  converge uniformément vers  $f$  sur tout segment  $[0, t]$  avec  $0 \leq t < 1$ . On en déduit alors par un argument de double limite que  $f^n(a_n) \rightarrow f(a)$ . Ainsi,  $f(a) = f(1/2)$  puis par stricte croissance de  $f$ ,  $a = 1/2$ . La convergence presque sûr de  $M_n$  entraîne alors la convergence presque sûr de  $Z_n$ .

**Lemme 1.** *La suite des fonctions génératrices  $f^n$  converge simplement vers une fonction  $f$  sur  $[0, 1)$ .*

*Preuve du lemme.* On montre que la suite de mesures  $g_n$  converge vaguement vers une mesure  $g$ . Comme les fonctions  $x \mapsto s^x \mathbf{1}_{x \geq 0}$  sont bornés sur  $\mathbf{R}$  pour  $s \in [0, 1)$ , la convergence vague assure la convergence des fonctions génératrices sur  $[0, 1)$ . On rappelle en annexe les principaux résultats sur la convergence de mesures.

On commence par exclure le cas où la masse de  $g$  se réduit en 0 c'est à dire, comme  $g_n[0] \rightarrow q$ , où  $g_n \xrightarrow{v} q\delta_0$ . Soit alors  $n_k$  une extraction tel que  $g_{n_k}$  converge vaguement vers une mesure  $g \neq q\delta_0$ . En s'appuyant sur cette sous-suite, nous allons montrer que  $g_n$  converge vaguement vers  $g$ . Pour cela, on encadre  $n$  par deux termes adjacents de l'extraction

$$u_n = n_k \leq n < n_{k+1} = v_n$$

de sorte que

$$g_n = f_{0:n} = f_{0:u_n} \circ f_{u_n:n}.$$

Il s'agit donc de montrer que  $f_{u_n:n}$  converge vaguement vers  $\delta_1$ . Pour cela, donnons nous une extraction  $n'$  tel que  $f_{u_{n'}:n'}$  converge vaguement vers une mesure  $h_1$ . Quitte à extraire de nouveau on peut supposer que  $f_{n':v_{n'}}$  converge également, vaguement, vers une mesure  $h_2$ . Comme rappelé

précédemment, la converge vague entre la convergence des fonctions génératrices sur  $[0, 1)$ . Partant de l'égalité

$$g_{v_{n'}}(s) = f_{0:u_{n'}} \circ f_{u_{n'}:n'} \circ f_{n':v_{n'}}$$

nous avons donc pour tout  $s \in [0, 1)$ ,

$$g(s) = g(h_1(h_2(s))).$$

Ayant exclus le cas où  $g$  est constant,  $g$  est strictement croissante donc  $h_1(h_2(s)) = s$  pour tout  $s \in [0, 1)$  ce qui sur les fonctions génératrices impliquent que  $h_1(s) = h_2(s) = 1$ . Par un argument de compacité usuel, cela garantit la convergence de  $f_{u_n:n}$  vers  $\delta_1$ .  $\square$

### Critère asymptotiquement dégénérée - Théorème 26

La réciproque est relativement direct. Pour tout  $z \geq 1$  avec  $\mathbf{P}(Z_0 = z) > 0$ ,

$$\mathbf{P}(Z_\infty = 0) \geq \mathbf{P}(Z_n = z, \forall n \geq 0) = \mathbf{P}(Z_0 = z) \left( \prod_{n=1}^{\infty} f_n[1] \right)^z > 0 \quad (91)$$

car la convergence de la somme  $\sum_{n=1}^{\infty} (1 - f_n[1])$  vers une valeur fini est équivalente à la convergence du produit  $\prod_{n=1}^{\infty} f_n[1]$  vers une valeur non nul<sup>10</sup>. On montre le sens direct en utilisant le lemme de Borel-Cantelli. Supposons que  $\sum_{n=1}^{\infty} (1 - f_n[1]) = \infty$ . Dire que  $Z_\infty = z$  c'est qu'à partir d'un certain rang,  $Z_n = z$  et ainsi les évènements  $A_n = \{\sum_{i=1}^z X_{i,n} = z\}$  se réalise tous à partir d'un certain rang. En passant à cette formulation, nous avons gagné l'indépendance des évènements. Il s'agit donc de montrer que  $\sum_{n=1}^{\infty} \mathbf{P}(A_n^c) = \infty$ . Pour que  $A_n$  est peu de chance de se réaliser, il y a essentiellement trois cas : 1. on a pas d'enfants avec une probabilité trop grande 2. on a plus de deux enfants avec une probabilité trop grande 3. on a un enfant avec une probabilité suffisante mais pas assez pour que tous ait un enfant. Pour distinguer ces cas et les traiter de la manière approprié on raisonne suivant que  $f_n[1]$  est supérieur à  $1/3$  ou non.

- si  $f_n[1] \geq 1/3$  alors on est dans le cas 3 et on utilise la majoration suivante

$$\mathbf{P}(A_n^c) \geq \mathbf{P}(X_{1,n} \neq 1, X_{2,n} = 1, \dots, X_{z,n} = 1) \geq 3^{-z-1}(1 - f_n[1]). \quad (92)$$

- si  $f_n[1] \leq 1/3$  alors on est dans le cas 1 ou 2 c'est à dire que  $f_n[0] \geq 1/3$  (cas 1) ou  $f_n[> 1] \geq 1/3$  (cas 2). Dans le premier cas,

$$\mathbf{P}(A_n^c) \geq \mathbf{P}(X_{1,n} = X_{2,n} = \dots = X_{z,n} = 0) \geq 3^{-z}. \quad (93)$$

Dans le second cas,

$$\mathbf{P}(A_n^c) \geq \mathbf{P}(X_{1,n}, X_{2,n}, \dots, X_{z,n} > 1) \geq 3^{-z}. \quad (94)$$

Finalement, dans tous les cas

$$\mathbf{P}(A_n^c) \geq 3^{-z}(1 - f_n[1]). \quad (95)$$

### Limite de Yaglom dans le cas variable

Rappelons que dans le cas classique, on a montré la convergence du processus convenablement renormalisé vers une loi exponentielle en regardant la transformée de Laplace, et en utilisant un résultat de convergence uniforme. Cette convergence uniforme correspond avec nos nouvelles notions, à la convergence uniforme de la fonction de forme normalisé par l'espérance du processus conditionnelle vers 1. C'est en généralisant ce résultat aux fonctions de formes de processus à environnement variable que nous allons établir la convergence vers la limite de Yaglom.

10. En passant au log et par sommation des équivalents avec  $\log(x) \sim 1 - x$  lorsque  $x \rightarrow 1$ .

**Proposition 31.** *Sous l'hypothèse B,  $\frac{1}{\varphi^n(1)}\varphi^n$  converge uniformément vers 1 sur  $[0, 1]$  c'est à dire*

$$\sup_{0 \leq s \leq 1} |\varphi^n(s) - \varphi^n(1)| = o(\varphi^n(1)).$$

Observons en particulier que  $\varphi^n(0) \sim \varphi^n(1)$  d'où

$$\mathbf{E}[Z_n | Z_n > 0] = \frac{\mu_n}{\mathbf{P}(Z_n > 0)} = 1 + \mu_n \varphi_n(0) \sim \mu_n \varphi_n(1) = \frac{\mu_n}{2} \sum_{k=1}^n \frac{\nu_k}{\mu^{k-1}}. \quad (96)$$

Le premier équivalent est dû au fait que l'on est en régime critique *i.e.*  $\mathbf{E}[Z_n | Z_n > 0] \rightarrow \infty$ . Notons qu'on a alors l'estimation

$$\frac{1}{\mu_n} = o(\varphi^n(1)) = o\left(\sum_{k=1}^n \frac{\nu_k}{\mu^{k-1}}\right). \quad (97)$$

Ainsi, la normalisation par  $\varphi^n(1)$  est équivalente à la normalisation par la moyenne du processus. En vu de la démonstration de ce résultat, rappelons l'expression de la fonction de forme

$$\varphi^n(s) = \sum_{k=1}^n \frac{\varphi_k(f_{k:n}(s))}{\mu^{k-1}}.$$

La convergence uniforme de cette somme se fait via deux mécanismes qui enclenchent la convergence pour séparément le début et la fin de la somme. D'une part, dans les premiers termes, la différence  $n-k$  est grande de sorte que  $f_{k:n}(s)$  contient un grand nombre d'itérés. Ces itérations font que l'on se rapproche de 1 et donc que  $\varphi_k(f_{k:n}(s))$  se rapproche de  $\varphi_k(1)$ . D'autre part, les derniers termes de la somme, disons supérieurs à  $r+1$ , vont être contrôlée par  $\frac{1}{\mu^r}$  qui est négligeable devant  $\varphi^n(1)$ . Le jeu va alors de bien choisir l'entier  $r$  et de séparer la somme selon  $r$  en traitant chaque partie suivant l'idée exposée précédemment.

Commençons par la première partie. L'ingrédient fondamental est d'avoir un contrôle de l'écart de  $\varphi_k$  à sa valeur en 1, au voisinage de 1. Dans son sens, on énonce sans démonstration le résultat suivant que l'on trouve dans [Ker20] (lemme Z) et qui repose sur une étude fine de la dérivée de la fonction de forme.

**Lemme 2.** *Pour tout  $a \in \mathbf{N}_0$ ,*

$$\sup_{s \leq t \leq 1} |\varphi_k(t) - \varphi_k(1)| \leq 2\mu_k \nu_k^2(1-s) + 2a\nu_k(1-s) + \frac{2}{\mu_k^2} \mathbf{E}[X_k^2 : X_k > a]. \quad (98)$$

Il nous faut un contrôler en  $\varepsilon \nu_k$ . Pour cela, on a une certaine liberté sur  $s$  et  $a$ . On prend  $s$  sous la forme

$$s_k = 1 - \frac{\varepsilon}{1 + \mu_k}$$

Pour ce choix

$$\sup_{s \leq t \leq 1} |\varphi_k(t) - \varphi_k(1)| \leq \frac{2\mu_k \nu_k^2 \varepsilon}{1 + \mu_k} + \frac{2a\nu_k \varepsilon}{1 + \mu_k} + \frac{2}{\mu_k^2} \mathbf{E}[X_k^2 : X_k > a] \quad (99)$$

et l'on peut faire tomber une dépendance en  $\nu_k$  dans le premier terme en utilisant l'hypothèse ??,  $\mu_k \nu_k \leq c_A(1 + \mu_k)$  d'où

$$\frac{2\mu_k \nu_k^2 \varepsilon}{1 + \mu_k} \leq 2c_A \nu_k \varepsilon. \quad (100)$$

Intéressons maintenant au troisième terme. Il faut avoir un contrôle en  $\varepsilon \nu_k$  pour un bon choix de  $a$ . C'est alors que nous avons besoin de l'hypothèse B. En choisissant  $a = c_\varepsilon(1 + \mu_k)$  on trouve

$$\frac{2}{\mu_k^2} \mathbf{E}[X_k^2 : X_k > a] \leq 4\varepsilon \nu_k. \quad (101)$$

Et le deuxième terme devient

$$\frac{2a v_k \varepsilon}{1 + \mu_k} = 2c_\varepsilon v_k. \quad (102)$$

Finalement en combinant le tout,

$$\sup_{s_k \leq t \leq 1} |\varphi_k(t) - \varphi_k(1)| \leq C v_k \varepsilon \quad (103)$$

où  $C = 2c_A + 2c_\varepsilon + 4$ . Avec notre estimation précédente, on introduit naturellement l'entier de séparation

$$r = \max\{k \leq n \mid f_{k:n}(0) \geq s_k\}.$$

Alors,

$$\left| \sum_{k=1}^r \frac{\varphi_k(f_{k:n}(s))}{\mu^{k-1}} - \sum_{k=1}^r \frac{\varphi_k(1)}{\mu^{k-1}} \right| \leq C \varepsilon \sum_{k=1}^r \frac{v_k}{\mu^{k-1}}. \quad (104)$$

Passons au deuxième morceau. Dans ce cas, la moyenne nous donnera le facteur négligeable, on majore donc simplement la fonction de forme par sa valeur en 1,

$$\left| \sum_{k=r+1}^n \frac{\varphi_k(f_{k:n}(s))}{\mu^{k-1}} - \sum_{k=r+1}^n \frac{\varphi_k(1)}{\mu^{k-1}} \right| \leq 3 \sum_{k=r+1}^n \frac{\varphi_k(1)}{\mu^{k-1}}. \quad (105)$$

On reconnait à un  $\mu_r$  près la fonction de forme de  $f_{r:n}$  en 1. On peut se ramener en 0 selon l'hypothèse A

$$\sum_{k=r+1}^n \frac{\varphi_k(1)}{\mu^{k-1}} \leq 2c_A \sum_{k=r+1}^n \frac{\varphi_k(f_{k:n}(0))}{\mu^{k-1}} \quad (106)$$

or,

$$\sum_{k=r+1}^n \frac{\varphi_k(f_{k:n}(0))}{\mu^{k-1}} = \frac{1}{\mu_r} \varphi_{r:n}(0) = \frac{1}{\mu^r(1-f_{r:n}(0))} - \frac{1}{\mu^n} \leq \frac{1}{\mu^r(1-f_{r:n}(0))}. \quad (107)$$

On aimerait se ramener à  $f_{r+1:n}$  afin d'avoir une minoration en  $s_{r+1}$ . On écrit

$$\frac{1}{\mu^r(1-f_{r:n}(0))} = \frac{1}{\mu^r} \left\{ \frac{1}{1-f_{r+1:n}(0)} + \varphi_r(f_{r+1:n}(0)) \right\}. \quad (108)$$

Le second terme se contrôle facilement en

$$\varphi_r(f_{r+1:n}(0)) \leq 2\varphi_r(1) = 2c_A(1 + \mu_r). \quad (109)$$

Et maintenant le premier terme peut se contrôler en

$$\frac{1}{1-f_{r+1:n}(0)} \leq \frac{1}{1-s_{r+1}} = \frac{1 + \mu_{r+1}}{\varepsilon}. \quad (110)$$

En combinant le tout on obtient

$$\frac{1}{\mu^r(1-f_{r:n}(0))} \leq \left\{ 2c_A + \frac{1}{\varepsilon} \right\} \left( \frac{1}{\mu^r} + \frac{1}{\mu^{r+1}} \right). \quad (111)$$

Et ainsi,

$$\sum_{k=r+1}^n \frac{\varphi_k(1)}{\mu^{k-1}} \leq \frac{C}{\varepsilon} \left( \frac{1}{\mu^r} + \frac{1}{\mu^{r+1}} \right). \quad (112)$$

Maintenant (97) assure qu'il existe un entier  $r_\varepsilon$  tel que pour  $r \leq r_\varepsilon \leq n$ ,

$$\left( \frac{1}{\mu^r} + \frac{1}{\mu^{r+1}} \right) \leq \varepsilon^2 \sum_{k=1}^{r+1} \frac{v_k}{\mu^{k-1}} \leq \varepsilon^2 \sum_{k=1}^n \frac{v_k}{\mu^{k-1}}. \quad (113)$$

Mais d'après 96

$$\frac{1}{\mathbf{P}(Z_n > 0)} \sim \frac{1}{2} \sum_{k=1}^n \frac{v_k}{\mu^{k-1}} \quad (114)$$

donc comme par hypothèse  $q = 1$ ,

$$\sum_{k=1}^n \frac{v_k}{\mu^{k-1}} \longrightarrow \infty. \quad (115)$$

L'inégalité reste ainsi valable pour  $r \leq r_\varepsilon$  pourvu que  $n$  soit assez grand. Alors,

$$\sum_{k=r+1}^n \frac{\varphi_k(1)}{\mu^{k-1}} \leq C\varepsilon \sum_{k=1}^n \frac{v_k}{\mu^{k-1}}. \quad (116)$$

Ce qui conclut la démonstration.

### Vitesse de convergence dans le cas variable

La construction du processus biaisé se généralise sans soucis au cas d'un environnement variable. La suite de variable aléatoire

$$W_n = W_n^f = \frac{Z_n^f}{\mu^n}$$

est toujours une  $\mathcal{F}_n$ -martingale donc la suite de marginales

$$\mathbf{GW}_{n,f}^b = W_n \cdot \mathbf{GW}_{n,f}$$

vérifie la condition de compatibilité donc se prolonge en une mesure  $\mathbf{GW}_f^b$  sur  $\mathcal{T}$ . Nous noterons  $Z_n^b$  la variable aléatoire  $Z_n$  sous  $\mathbf{GW}_f^b$ . Comme dans le cas classique, on peut écrire

$$Z_n = 1 + \sum_{j=1}^n Z_{n,j}$$

où  $Z_{n,j}$  est un processus de Galton-Watson en environnement aléatoire  $\mathbf{f}_{\geq j} = (f_n)_{n \geq j}$  de population initiale  $I_j$  qui à la loi du biais de  $f_j$ . Et ensuite  $Z_{n,j} = L_{n,j} + R_{n,j}$  d'où  $Z_n = L_n + R_n$  avec,

$$L_n = \sum_{j=1}^n L_{n,j}, \quad R_n = 1 + \sum_{j=1}^n R_{n,j}.$$

Et l'on a alors,

$$Z_n \mid Z_n > 0 \stackrel{\text{Loi}}{=} Z_n^b \mid A_n \quad (117)$$

où  $A_n = \{L_n = 0\} = \bigcap_{j=1}^n A_{n,j}$  avec  $A_{n,j} = \{L_{n,j} = 0\}$ . On introduit également par la même construction la variable aléatoire  $\tilde{R}_n = 1 + \sum_{j=1}^n \tilde{R}_{n,j}$  de loi  $Z_n^b \mid A_n$  où

$$\tilde{R}_{n,j} = \mathbf{1}_{A_{n,j}} R_{n,j} + \mathbf{1}_{A_{n,j}^c} R'_{n,j}$$

où  $R'_{n,j} \sim Z_{n,j}^b \mid A_{n,j}$ . Ces variables aléatoires ont la même interprétation que dans le cas classique. Le point de départ est alors le même, en utilisant la méthode de Stein, et en piochant nos variables aléatoires dans les variables précédemment construites on obtient la formule suivant, généralisation de l'équation (35) à un environnement variable

$$d_W(\bar{Y}_n, \text{Exp}) \leq \frac{2}{b_n} \left\{ \frac{1}{2} + \mathbf{E}[|\tilde{R}_n - R_n|] \right\}. \quad (118)$$

Il s'agit donc d'estimer l'écart entre  $\tilde{R}_n$  et  $R_n$ . On se ramène naturellement à contrôler l'écart entre  $R'_{n,j}$  et  $R_{n,j}$  et ces deux variables aléatoires étant égales sur  $A_{n,j}$ , on peut se restreindre à  $A_{n,j}^c$  et traiter les deux variables séparément. On a ainsi,

$$\mathbf{E}(|R_n - \tilde{R}_n|) \leq \sum_{j=1}^n \mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] + \mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}]. \quad (119)$$

Nous allons reprendre la démarche de nos calculs en veillant au fait que nous sommes maintenant en environnement variable. Ainsi, nous devons garder trace de notre environnement, ce dernier étant amené à être shifter. De manière général, nous ferons figurer un exposant de la forme  $\geq j$  pour signifier que l'on regarde notre processus dans l'environnement  $f_j = (f_n)_{n \geq j}$ . Cette notation fera sens quand nous l'employerons.

Commençons par le terme en  $\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}]$ . Le début reste le même, on a toujours

$$\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \mathbf{E}[\mathbf{E}[Z_{n,j} | I_j] \mathbf{P}(Z_{n,j} > 0 | I_j)]. \quad (120)$$

Maintenant il faut prendre garde à l'environnement.  $Z_{n,j}$  est un processus de Galton-Watson de population initiale  $I_j$  dans l'environnement  $f_{\geq j}$ . Ainsi,

$$\mathbf{E}[Z_{n,j} | I_j] = I_j \mathbf{E}[Z_{n-j}^{f_{\geq j}}] = I_j \frac{\mu^n}{\mu^j} \quad (121)$$

et  $\mathbf{P}(Z_{n,j} > 0 | I_j) = I_j \mathbf{P}(Z_{n-j}^{\geq j} > 0)$ . En réinjectant nous obtenons

$$\mathbf{E}[R_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \mathbf{E}(I_j^2) \mathbf{E}[Z_{n-j}^{f_{\geq j}}] \mathbf{P}(Z_{n-j}^{\geq j} > 0) \quad (122)$$

avec,

$$\mathbf{E}(I_j^2) = \frac{\mathbf{E}((X_j - 1)^2 X_j)}{\mathbf{E}(X_j)} = \frac{f_j'''(1) + f_j''(1)}{f_j'(1)}. \quad (123)$$

Passons au deuxième terme  $\mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}]$ . Le début est le même, nous avons toujours

$$\mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \mathbf{E}[Z_{n,j}^b] \mathbf{P}(A_{n,j}^c). \quad (124)$$

Dorénavant  $\mathbf{E}[Z_{n,j}^b] = \frac{\mu^n}{\mu^j} \mathbf{E}(I_j)$  mais l'on a encore

$$\mathbf{P}(A_{n,j}^c) = \mathbf{E}(I_j) \mathbf{P}(Z_{n-j}^{\geq j} > 0) \quad (125)$$

avec,

$$\mathbf{E}(I_j) = \frac{\mathbf{E}(X_j(X_j - 1))}{\mathbf{E}(X_j)} = \frac{f_j''(1)}{f_j'(1)} \quad (126)$$

Ainsi,

$$\mathbf{E}[R'_{n,j} \mathbf{1}_{A_{n,j}^c}] \leq \frac{\mu^n}{\mu^j} \left( \frac{f_j''(1)}{f_j'(1)} \right)^2 \mathbf{P}(Z_{n-j}^{\geq j} > 0). \quad (127)$$

En regroupons les estimations nous avons alors

$$\mathbf{E}(|R_n - \tilde{R}_n|) \leq \mu^n \sum_{j=1}^n \frac{1}{\mu^j} \left( \frac{f_j'''(1) + f_j''(1)}{f_j'(1)} + \left( \frac{f_j''(1)}{f_j'(1)} \right)^2 \right) \mathbf{P}(Z_{n-j}^{\geq j} > 0) \quad (128)$$

Le terme entre parenthèse est une expression faisant intervenir les moments factoriels de nos environnement. A l'aide de l'hypothèse C on peut simplifier cette expression. D'abord avec (??)

$$\frac{f_j'''(1) + f_j''(1)}{f_j'(1)} + \left( \frac{f_j''(1)}{f_j'(1)} \right)^2 \leq \frac{f_j''(1)}{f_j'(1)} \left\{ 1 + c(1 + f_j'(1)) + \frac{f_j''(1)}{f_j'(1)} \right\} \quad (129)$$

ensuite avec l'inégalité

$$\frac{f_j''(1)}{f_j'(1)} \leq 4c_{1/2}(1 + f_j'(1)) \quad (130)$$

qui découle de l'hypothèse B, nous en déduisons l'existence d'une constante  $C$  tel que

$$\mathbf{E}(|R_n - \tilde{R}_n|) \leq C \mu^n \sum_{j=1}^n \frac{\nu_j}{\mu^{j-1}} (1 + \mu_j) \mathbf{P}(Z_{n-j}^{\geq j} > 0). \quad (131)$$

Il reste à estimer  $\mathbf{P}(Z_{n-j}^{\geq j} > 0)$ . On reprend pour cela la démarche de la partie précédente en passant par les fonctions de formes. On écrit

$$\mathbf{P}(Z_{n-j}^{\geq j} > 0) = 1 - f_{j:n}(0) = \left( \frac{\mu^j}{\mu^n} + \varphi_{j:n}(0) \right)^{-1} \leq \left( \frac{c\mu_j}{2} \varphi_{j:n}(1) \right)^{-1} \leq \frac{C}{\mu_j(\nu^n - \nu^j)}. \quad (132)$$

En injectant cela dans l'équation précédente, nous obtenons bien le résultat voulu :

$$d_W\left(\frac{1}{b_n} Z_n \mid Z_n > 0, \text{Exp}\right) \leq C \left( \frac{1}{\mu^n \nu^n} + \frac{1}{\nu^n} \sum_{j=1}^{n-1} \frac{\nu_j}{\mu^{j-1}} \frac{1 + \mu_j}{\mu^j(\rho^n - \rho^j)} + \frac{\nu_n}{\mu^{n-1}} (1 + \mu_n) \right). \quad (133)$$

## Références

- [CTJP23] Natalia Cardona-Tobón, Arturo Jaramillo, and Sandra Palau. Rates on yaglom's limit for galton-watson processes in varying environment, 2023.
- [Ker20] Götz Kersting. A unifying approach to branching processes in a varying environment. *Journal of Applied Probability*, 57 :196–220, 03 2020.
- [KV17] Götz Kersting and Vladimir Vatutin. *Discrete Time Branching Processes in Random Environment*. 11 2017.
- [LPP95] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of  $l \log l$  criteria for mean behavior of branching processes. *The Annals of Probability*, 23(3) :1125–1138, 1995.
- [MS08] I. MaCphee and H.-J Schuh. A galton-watson branching process in varying environments with essentially constant offspring means and two rates of growth1. *Australian Journal of Statistics*, 25 :329 – 338, 02 2008.
- [PR11] Erol Peköz and Adrian Röllin. New rates for exponential approximation and the theorems of rényi and yaglom. *The Annals of Probability*, 39, 03 2011.
- [Ros11] Nathan Ross. Fundamentals of stein's method, 2011.