Chapter 1

Introduction to Galton-Watson branching processes

1.1 Some historical background of the model

The model first arouse from F. Galton's statistical investigation of the extinction of family names. In the nineteenth century, there was concern amongst the Victorians that aristocratic surnames were becoming extinct. In 1873, Galton originally posed the question regarding the probability of such an event in an issue of "The Educational Times", and the Reverend H. W. Watson later replied with a solution. Together, they then wrote a paper in 1874 entitled "On the probability of the extinction of families" for the Journal of the Royal Anthropological Institute.

Since then, this model has found applications in biology (fixation of genes, early evolution of bacterial colonies), chemistry (chemical chain reactions) and physics (cosmic rays). Unfortunately, this model is in fact of limited applicability in understanding the actual family name distributions, since family names may change for many other reasons than dying out of male family line.

1.2 Definition of (single-type) Galton-Watson branching process

The Galton-Watson branching process (or GW-process for short) is the simplest possible model for a population evolving in time. It is based on the assumption that individuals in the population give birth to a number of children independently of each other and all with the same distribution. More precisely, the model can be described as follows:

- We start the process with a single individual, the *zero generation* of the population.
- This individual gives birth to a random number $X \in \mathbb{N}_0$ of children, with $\mathbb{E}(X) \in (0, +\infty)$. These children constitute the *first generation* of the population.
- Each of the individuals in this first generation have children of their own, all of them with the same distribution as X and independently of all the other individuals in the population. These constitute the *second generation* of the population. This generation in turn gives rise to a *third generation* of individuals by the same rules as the previous generation, and so on.

Notice that, if Z_n denotes the number of individuals in the n-th generation ($n \in \mathbb{N}_0$), then Z_n satisfies the recurrence relation

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^{(n)}$$

where:

- $Z_0 \equiv 1$ by convention,
- $\{X_i^{(n)} : i \in \mathbb{N}\,,\, n \in \mathbb{N}_0\}$ is an array of i.i.d. random variables with $X_i^{(n)} \sim X$ for all i, n.

The sequence $(Z_n)_{n\in\mathbb{N}_0}$ is what is typically known as the GW-process with offspring distribution X. However, sometimes by a GW-process we shall understand the entire genealogical tree induced by the population, i.e the collection $\{(n, i, X_i^{(n)}) : n \in \mathbb{N}_0\,,\, i \in \{1, \ldots, Z_n\}\}$, and not just $(Z_n)_{n\in\mathbb{N}_0}$. When is it that we mean one or the other will always be clear from the context.

1.3 Survival vs Extinction

Consider the *extinction event* in which the population eventually dies out, i.e.

$$\{\text{extinction}\} = \{\exists\, n \in \mathbb{N} : Z_n = 0\} = \bigcup_{n\in\mathbb{N}} \{Z_n = 0\},$$

and define η to be the *extinction probability*, i.e.

$$\eta := P(extinction) = P\left(\bigcup_{n\in\mathbb{N}} \{Z_n = 0\}\right).$$

It is easy to see that $\eta = 0$ whenever $P(X = 0) = 0$, since in this case each generation must have at least one individual. On the other hand, if $P(X = 0) > 0$ then η is always strictly positive. Indeed, this can happen for example if the original individual has 0 children, which gives the bound

$$\eta \geq P(X = 0) > 0.$$

Formulated in these terms, Galton's goal was to determine the value of η. In his solution provided to Galton, Watson reached the conclusion that extinction was **always** the case if $P(X = 0) > 0$, i.e. that $\eta = 1$. However, Watson's original derivation had a small error in it and, as it turns out, his conclusion was erroneous: in some cases it is truly possible to have $\eta \in (0, 1)$ and thus to get a positive probability that the population carries on for an infinite number of nonzero generations. When is it that we have $\eta = 1$ or $\eta < 1$ will depend, as a matter of fact, on the value of $\mathbb{E}(X)$. Indeed, we will prove the following.

Theorem 1.3.1. *Suppose that $\mu := \mathbb{E}(X) \in (0, +\infty)$. Then:*

*i. If $\mu < 1$ then $\eta = 1$, i.e. the population **always** dies out.*

ii. If $\mu > 1$ then $\eta < 1$, i.e. there is a positive probability of survival.

iii. If $\mu = 1$ then $\eta = 1$ unless $P(X = 1) = 1$, in which case $\eta = 0$.

Moreover, if $G_X : [0,1] \to [0,1]$ *denotes the probability generating function of* X *defined as*

$$G_X(s) := \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k P(X = k),$$

then η *is the smallest solution in* $[0,1]$ *of the equation*

$$\eta = G_X(\eta). \tag{1.1}$$

In fact, if $P(X = 1) < 1$ *then any solution* $\psi \in [0,1]$ *of* (1.1) *satisfies* $\psi \in \{\eta, 1\}$.

Remark 1.3.2. *To define* G_X *we use the convention* $0^0 = 1$, *so that* $G_X(0) = P(X = 0)$.

1.4 Proof of Theorem 1.3.1 - Step I

We begin by stating some elementary properties of G_X in the following proposition, whose proof is left as an exercise.

Proposition 1.4.1. *The function* G_X *is continuous on* $[0,1]$ *and twice differentiable on* $(0,1)$. *Moreover, for* $s \in (0,1)$ *we have*

$$G_X'(s) = \mathbb{E}(X s^{X-1}) \qquad \text{and} \qquad G_X''(s) = \mathbb{E}(X(X-1)s^{X-2}),$$

In particular, G_X *is increasing and convex. Also,* G_X *is differentiable from the left at* $s = 1$ *and its derivative satisfies* $G_X'(1) = \mathbb{E}(X)$.

Now, the first step of the proof is to establish the following:

Lemma 1.4.2. *Both* η *and* 1 *are solutions of the equation* (1.1) *and any other solution* $\psi \in [0,1]$ *satisfies* $\eta \le \psi \le 1$.

Proof. That 1 is a solution of (1.1) is immediate to see. Let us check that η is also a solution.

For each $n \in \mathbb{N}$ let us define

$$\eta_n := P(Z_n = 0).$$

Observe that, since $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$ for all n, by the continuity of probability measures we have that

$$\eta = \lim_{n \to +\infty} \eta_n. \tag{1.2}$$

Now, let

$$G_n(s) := \mathbb{E}(s^{Z_n})$$

be the probability generating function of the n-th generation. Notice that, by conditioning on Z_1, we have

$$G_{n+1}(s) = \mathbb{E}(s^{Z_{n+1}}) = \sum_{k=0}^{\infty} \mathbb{E}(s^{Z_{n+1}} | Z_1 = k) P(Z_1 = k) = \sum_{k=0}^{\infty} G_n^k(s) P(X = k) = G_X(G_n(s)).$$

Indeed, the second equality is just by definition of conditional expectation and the third is because, conditionally on $Z_1 = k$, we can write

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

where $\{Z^{(i)} : i = 1, \dots, k\}$ are independent GW-processes with offspring distribution X which are independent of Z_1.

In particular, if we evaluate this equality at $s = 0$, by Remark 1.3.2 we obtain the recurrence

$$\eta_{n+1} = G_X(\eta_n).$$

Upon taking the limit $n \to +\infty$, (1.2) together with the continuity of G_X yields that

$$\eta = G_X(\eta),$$

thus showing (1.1).

Finally, let $\psi \in [0, 1]$ be a solution of (1.1). Since clearly $\psi \leq 1$, it only remains to show that $\eta \leq \psi$. By (1.2), it will suffice to show that

$$\eta_n \leq \psi$$

for all $n \in \mathbb{N}$. We do this by induction. First, observe that by the monotonicity of G_X we have

$$\eta_1 = G_1(0) = G_X(0) \leq G_X(\psi) = \psi,$$

so that $\eta_1 \leq \psi$. Now, if $\eta_n \leq \psi$ for some $n \in \mathbb{N}$ then by the recurrence relation $G_{n+1} = G_X \circ G_n$ we get

$$\eta_{n+1} = G_{n+1}(0) = G_X(G_n(0)) = G_X(\eta_n) \leq G_X(\psi) = \psi.$$

Thus, this shows that $\eta_n \leq \psi$ for all n which, upon taking the limit $n \to +\infty$, yields that $\eta \leq \psi$. □

1.5 Proof of Theorem 1.3.1 - Step II

Now, we split the proof into three separate cases.

1.5.1 Case I : $P(X = 1) = 1$

Here we have $Z_n \equiv 1$ for all n, which immediately implies that $\eta = 0$. In addition, let us note that in this case $G_X(s) = s$ so that **every** $s \in [0, 1]$ is a solution of (1.1).

1.5.2 Case II : $P(X \leq 1) = 1$ but $P(X = 0) > 0$

If $p := P(X = 0) > 0$ then it is easy to see that $\eta_n = P(Z_n = 0) = 1 - (1 - p)^n \to 1$, so that by (1.2) we get $\eta = 1$. Being η the smallest solution of (1.1) by Lemma 1.4.2, it follows that $\eta = 1$ is the unique solution.

1.5.3 Case III : $P(X \geq 2) > 0$

. Observe that in this case one has $G''_X > 0$ by Proposition 1.4.1, so that G_X is **strictly** convex. It follows from this that (1.1) can have at most two solutions and it will have exactly two if and only if $\eta < 1$.

Now, define $f(s) := G_X(s) - s$ and observe that solutions of (1.1) are exactly the zeros of f. Suppose that $\eta < 1$. Then, since $f(\eta) = f(1) = 0$, by the mean value theorem there exists $c \in (\eta, 1)$ so that $f'(c) = 0$, i.e. $G'_X(c) = 1$. But since $G''_X > 0$ on $(0, 1)$, it follows that $G'_X(1) > G'_X(c) = 1$. Hence, since $\mu = G'_X(1)$ by Proposition 1.4.1, we conclude that if $\mu \leq 1$ then one must have $\eta = 1$.

On the other hand, if $\mu > 1$ then $f'(1) > 0$ and therefore there exists $\bar{s} \in (0, 1)$ (sufficiently close to 1) such that $f(\bar{s}) < f(1) = 0$. Thus, since $f(0) = G_X(0) = P(X = 0)$, one of the following two options must occur:

i. $f(0) = 0$, which immediately yields that $\eta = 0 < 1$,

ii. $f(0) > 0$, so that by continuity of f there exists $\hat{s} \in (0, \bar{s})$ with $f(\hat{s}) = 0$. Hence, $\eta = \hat{s} < 1$.

This covers all possible cases and thus concludes the proof.

1.6 The three regimes of a Galton-Watson process

According to Theorem 1.3.1, we can subdivide the class of Galton-Watson processes into three categories/regimes:

i. *subcritical*, if $\mu < 1$

ii. *supercritical*, if $\mu > 1$

iii. *critical*, if $\mu = 1$.

As we have shown, the supercritical regime is the only one for which there is a positive probability of survival of the population. Even though both in (i) and (ii) the population eventually dies out, we shall see that, in fact, the population in both regimes behaves rather differently.

Our goal is to characterize the behavior of the Galton-Watson process in each of these regimes. For example, one could ask the following questions:

- In the supercritical regime, how does Z_n asymptotically grow on the event of survival?
- In the sub/critical regimes, what is the rate of decay of $P(Z_n > 0)$? What can we say about the extinction time $\tau := \min\{n \in \mathbb{N} : Z_n = 0\}$?
- In the sub/critical regimes, what can we say about the behavior of the population conditioned on the event $\{Z_n > 0\}$ for n large?

1.7 Exercises

1. Prove Proposition 1.4.1.

Chapter 2

Preliminaries

We begin our analysis with a few preliminary notions to be used in the sequel.

2.1 Conditional expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X \geq 0$ some random variable defined on it. In principle, X is allowed to take the value $+\infty$. Given a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, we say that a random variable Z is the *conditional expectation of* X *given* $\mathcal{G}$, and denote by $Z = \mathbb{E}(X|\mathcal{G})$, if it satisfies:

C1. Z is $\mathcal{G}$-measurable, i.e. $\{Z \in B\} \in \mathcal{G}$ for any Borel set $B \in \mathcal{B}(\mathbb{R})$.

C2. For any $A \in \mathcal{G}$,

$$\int_A X dP = \int_A Z dP.$$

It can be seen that such a random variable Z always exists and, furthermore, it is P-unique, i.e. if Z, Z' are any two random variables satisfying (C1-C2) then $P(Z = Z') = 1$, see Section 2.6. Moreover, taking $A = \Omega$ in (C2) yields the fundamental property of conditional expectation

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})),$$

which we shall use extensively throughout our analysis.

In the sequel, we will only need to compute $\mathbb{E}(X|\mathcal{G})$ whenever $\mathcal{G}$ is the σ-algebra generated by a collection $Y = (Y_j)_{j \in J}$ of random variables, i.e. the smallest σ-algebra $\sigma(Y)$ on Ω which contains all the events $\{Y_j \in B\}$ with $B \in \mathcal{B}(\mathbb{R})$ and $j \in J$. In this case, $\mathbb{E}(X|\mathcal{G})$ can be interpreted as the expected value of X given the values of the collection Y. To stress this fact, in the sequel we shall adopt the notation $\mathbb{E}(X|Y)$ instead of $\mathbb{E}(X|\sigma(Y))$.

The following two propositions, whose proofs are left as an exercise, show us how to compute the conditional expectation $\mathbb{E}(X|Y)$ in the few particular cases we shall need.

Proposition 2.1.1. *Suppose* $Y = (Y_1, \ldots, Y_n)$ *is a vector of discrete random variables* $Y_k : \Omega \to \mathbb{R}$. *Show that* $\mathbb{E}(X|Y) = g(Y)$, *where the function* $g : R_Y \to [0, +\infty]$ *is given by*

$$g(y) := \mathbb{E}(X|Y = y) := \frac{\mathbb{E}(X \mathbb{1}_{\{Y=y\}})}{P(Y = y)}. \tag{2.1}$$

and $R_Y := \{y = (y_1, \ldots, y_n) \in \mathbb{R}^n : P(Y = y) > 0\}$.

Proposition 2.1.2. *Suppose $Y = (Y_k)_{k\in\mathbb{N}}$ is an infinite sequence of random variables $Y_k : \Omega \to \mathbb{R}$ and X is of the form*

$$X = f(U, Y)$$

for a measurable $f \geq 0$ and U independent of Y. Show that $\mathbb{E}(X|Y) = g(Y)$, with $g : \mathbb{R}^{\mathbb{N}} \to [0, +\infty]$ given by

$$g(y) = \mathbb{E}(f(U, y)). \tag{2.2}$$

Notice that this formula obeys the intuition

$$g(y) = \mathbb{E}(f(U, Y)|Y = y) = \mathbb{E}(f(U, y)|Y = y) = \mathbb{E}(f(U, y))$$

motivated by the Substitution Principle and the independence of U and Y.

2.2 Martingales

Given a sequence $W = (W_n)_{n\in\mathbb{N}_0}$ of nonnegative random variables, we will say it is a *martingale* whenever it satisfies the following two properties:

M1. $\mathbb{E}(W_n) < +\infty$ for all $n \in \mathbb{N}_0$,

M2. $\mathbb{E}(W_{n+1}|W_0, \ldots, W_n) = W_n$ for all $n \in \mathbb{N}_0$.

We will say that $W = (W_n)_{n\in\mathbb{N}}$ is a *submartingale* if instead of (M2) above we have that

$$\mathbb{E}(W_{n+1}|W_0, \ldots, W_n) \geq W_n$$

holds for all $n \in \mathbb{N}_0$. In particular, any martingale is a submartingale.

Martingales are well-studied objects in probability theory. In particular, we have the following convergence result, whose proof is classical (but not elementary!).

Theorem 2.2.1. *Any nonnegative submartingale $(W_n)_{n\in\mathbb{N}}$ satisfying that $\sup_{n\in\mathbb{N}} \mathbb{E}(W_n) < +\infty$ converges almost surely to some limiting random variable W_∞, and this limit is almost surely finite. In particular, any nonnegative martingale converges almost surely (to an almost surely finite limit). Furthermore, we have that:*

1. $W_n \xrightarrow{L^p} W_\infty \iff \sup_{n\in\mathbb{N}} \mathbb{E}(|W_n|^p) < +\infty$, *whenever $p > 1$.*
2. $W_n \xrightarrow{L^1} W_\infty \iff (W_n)_{n\in\mathbb{N}}$ *is uniformly integrable (see Section 2.4).*

We shall also need the following related result.

Theorem 2.2.2. *Consider a probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be an increasing sequence of sub-σ-algebras of $\mathcal{F}$ which generate $\mathcal{F}$, i.e. $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}$ for every n and $\sigma(\mathcal{F}_n : n \in \mathbb{N}) = \mathcal{F}$. Then, for any random variable $X \geq 0$ with $\mathbb{E}(X) < +\infty$, the sequence $(X_n)_{n\in\mathbb{N}}$ given by*

$$X_n := \mathbb{E}(X|\mathcal{F}_n)$$

is a martingale which converges P-almost surely to X.

2.3 Tightness

A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ (taking values in $\mathbb{R}$) is said to be *tight* if for each $\varepsilon > 0$ there exists $M_\varepsilon > 0$ such that

$$\sup_{n\in\mathbb{N}} P(|X_n| > M_\varepsilon) \leq \varepsilon.$$

Tightness has the following relationship with convergence in distribution.

Theorem 2.3.1. *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ is tight if and only if for each subsequence $(X_{n_k})_{n\in\mathbb{N}}$ there exists a further subsequence $(X_{n_{k_j}})_{j\in\mathbb{N}}$ which converges in distribution.*

Furthermore, one also has the following characterization of convergence in distribution.

Proposition 2.3.2. *Let $\{X_n : n \in \mathbb{N} \cup \{\infty\}\}$ be a collection of real-valued random variables on $\mathbb{R}$. Then, the sequence $(X_n)_{n\in\mathbb{N}}$ converges in distribution to X_∞ if and only if for each subsequence $(X_{n_k})_{n\in\mathbb{N}}$ there exists a further subsequence $(X_{n_{k_j}})_{j\in\mathbb{N}}$ converging in distribution to X_∞.*

Together with Theorem 2.3.1, this yields the following strategy for proving that a given sequence $(X_n)_{n\in\mathbb{N}}$ of random variables in $\mathbb{R}$ converges in distribution:

i. Show that $(X_n)_{n\in\mathbb{N}}$ is tight.

ii. Show that any convergent subsequence has the same limit.

2.4 Uniform integrability

We say that a collection $(X_j)_{j\in J}$ of random variables is *uniformly integrable* if

$$\lim_{M\to+\infty} \left(\sup_{j\in J} \mathbb{E}(|X_j| \mathbb{1}_{\{|X_j|>M\}}) \right) = 0.$$

We have the following relation between uniform integrability and the uniform boundedness of the moments of $(X_j)_{j\in J}$.

Proposition 2.4.1. *For any $p > 1$ the following implications hold:*

$$\sup_{j\in J} \mathbb{E}(|X_j|^p) < +\infty \Longrightarrow (X_j)_{j\in J} \text{ uniformly integrable} \Longrightarrow \sup_{j\in J} \mathbb{E}(|X_j|) < +\infty.$$

None of the reverse implications necessarily hold.

Finally, we have the following crucial result on uniform integrability.

Proposition 2.4.2. *If $(X_n)_{n\in N}$ is uniformly integrable and converges in distribution to some X, then $\mathbb{E}(|X|) < +\infty$ and, moreover, one has that*

$$\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X).$$

2.5 Exercises

1. Verify that any (nonnegative) martingale $(W_n)_{n\in\mathbb{N}_0}$ has constant expectation. In other words, show that $\mathbb{E}(W_n) = \mathbb{E}(W_0)$ for all n.

2. Let $Z = (Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process and define the sequence $W = (W_n)_{n\in\mathbb{N}_0}$ by the formula
$$W_n = \frac{Z_n}{\mu^n}.$$
Show that W is nonnegative martingale with constant expectation 1.

3. Prove Theorem 2.3.1. Here are some tips:

   i. Notice that for each $x \in \mathbb{R}$ there exists a subsequence $(X_{n_k})_{k\in\mathbb{N}}$ of $(X_n)_{n\in\mathbb{N}}$ such that the limit $\lim_{k\to+\infty} F_{X_{n_k}}(x)$ exists, where $F_{X_{n_k}}$ denotes the distribution function of X_{n_k}.
   ii. By means of a diagonal argument, obtain a subsequence for which the latter convergence holds for all x in a countable dense subset of $\mathbb{R}$. Use this fact to construct a candidate for a limiting distribution function F.
   iii. Show that F can be taken to be increasing and right-continuous. Finally, use tightness to show that it must also verify
$$\lim_{x\to-\infty} F(x) = 0 \qquad \text{and} \qquad \lim_{x\to+\infty} F(x) = 1.$$

4. Prove Proposition 2.3.2.

2.6 Existence/Uniqueness of the conditional expectation

The existence and uniqueness of $\mathbb{E}(X|\mathcal{G})$ is a direct consequence of the Radon-Nikodym theorem from measure theory. Let us recall this result.

Theorem 2.6.1 (Radon-Nikodym). *Let ν and μ be two measures on some measurable space $(\Omega, \mathcal{F})$ and suppose that μ is σ-finite. If $\nu \ll \mu$ then there exists a nonnegative measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for all $A \in \mathcal{F}$,*
$$\nu(A) = \int_A f d\mu.$$
Moreover, f is μ-unique: any other function g with these characteristics satisfies $\mu(\{f \neq g\}) = 0$. The function f is called the Radon-Nikodym derivative of ν with respect to μ and is denoted by $\frac{d\nu}{d\mu}$.

Now, consider a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X \geq 0$ defined on $(\Omega, \mathcal{F})$. Given a measure μ on $(\Omega, \mathcal{F})$, define $\mu|_{\mathcal{G}}$ to be its restriction to $\mathcal{G}$ and also $X \cdot \mu$ as the measure
$$(X \cdot \mu)(A) := \int_A X d\mu. \tag{2.3}$$
Notice that, given a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, a random variable Z satisfies (C1-C2) in the definition of the conditional expectation $\mathbb{E}(X|\mathcal{G})$ if and only if it is a Radon-Nikodym derivative of $[X \cdot P]|_{\mathcal{G}}$ with respect to $P|_{\mathcal{G}}$. By Theorem 2.6.1, such a derivative exists and is also P-unique. From this, the existence and uniqueness of $\mathbb{E}(X|\mathcal{G})$ immediately follows.

Chapter 3

Size-biased Galton-Watson trees and branching processes with immigration

3.1 Size-biasing

Given a nonnegative random variable X satisfying $0 < \mathbb{E}(X) < +\infty$, a size-biasing of X will be any random variable $\widehat{X}$ whose distribution is given by the formula

$$P(\widehat{X} \in A) = \frac{\mathbb{E}(X \mathbb{1}_A(X))}{\mathbb{E}(X)}.$$

In the case that X is an $\mathbb{N}_0$-valued, $\widehat{X}$ is also $\mathbb{N}_0$-valued and its distribution is given by

$$P(\widehat{X} = k) = \frac{kP(X = k)}{\mu}$$

for all $k \in \mathbb{N}_0$. Note that $P(\hat{X} = 0)$, so that in fact $\hat{X} \geq 1$.

3.1.1 Probabilistic interpretation of size-biasing

Suppose that we have originally an urn with balls such that, whenever we reach into the urn and grab a ball uniformly at random, the probability of picking one with label $k \in \mathbb{N}_0$ is $P(X = k)$.[1] Suppose now we that take a new urn and fill it with balls so that for each ball with label k in the original urn we put k balls of label k in the new urn. Then, if we reach into the new urn and grab a ball at random, the probability of picking one which has label k is proportional to $kP(X = k)$ and thus must coincide with the size-biased probability $P(\widehat{X} = k)$.

3.2 Size-biased Galton-Watson trees

Note that any Galton-Watson process defines a (possibly infinite) rooted tree T in a natural way: for $i \leq Z_n$ the random variable $X_i^{(n)}$ records the number of descendants in T of the i-th individual from the n-th generation. In what follows, given a tree t we will denote by t_n the restriction of t to its first $n \in \mathbb{N}_0$ generations and by ∂t_n the set of individuals belonging to the n-th generation.

[1]For the sake of this interpretation, we assume that X is such that it is indeed possible to have such an urn.

For reasons that will be made clear later, we are interested in performing a size-biasing of T using $(Z_n)_{n\in\mathbb{N}_0}$ as labels. Informally, one could proceed for any finite n as follows:

i. Imagine an urn containing trees which only have up to n generations and define the label of any such tree t in the urn as $Z_n(t)$, the number of individuals in the n-th generation of t.

ii. Suppose that, for each tree t as above, the number of copies of t in the urn is proportional to the probability $P(T_n = t)$.[2] In particular, for each $k \in \mathbb{N}_0$, the number of trees in the urn with label k is proportional to $P(Z_n = k)$.

iii. Now, imagine that you take a new urn and fill it with all the vertices from the n-th generations of the trees from the original urn, and then you pick a vertex at random from this new urn. Given a vertex v in this new urn, set its label as $Z_n(t_v)$, where t_v is the tree in the first urn to which v belongs.

iv. Note that, for any $k \in \mathbb{N}_0$, the number of vertices in the new urn with label k is proportional to $kP(Z_n = k)$ by construction, since each tree with label k in the original urn surrenders exactly k vertices to the new urn. Thus, this is analogous to the description of size-biasing of Section 3.1.1. In particular, the probability of a vertex picked at random having label k is proportional to $kP(Z_n = k)$ and thus coincides with the size-biased probability $P(\widehat{Z}_n = k)$.

v. As a matter of fact, if v_n denotes the chosen vertex and $\widehat{T}_n := t_{v_n}$ the tree to which it belongs, then the construction above has the property that

$$P(\widehat{T}_n = t_n\,,\, v_n = v) = \frac{1}{\mu^n} P(T_n = t) \tag{3.1}$$

for any rooted tree t and $v \in \partial t_n$.

The idea is now to do this again, rigorously this time, but for the full Galton-Watson tree T. More precisely, we wish to construct a random tree $\widehat{T}$ with a distinguished path $(v_n)_{n\in\mathbb{N}}$ inside it such that, if $\widehat{T}_n$ denotes the restriction to $\widehat{T}$ to its first n generations, for each n the pair $(\widehat{T}_n, v_n)$ satisfies (3.1), i.e. we can think of $(\widehat{T}_n, v_n)$ as being obtained via the urn procedure detailed above. To construct the pair $(\widehat{T}, (v_n)_{n\in\mathbb{N}_0})$ we do as follows:

i. Start with an initial individual v_0, and let it have a random number of children $\widehat{X}_1 \sim \widehat{X}$.

ii. Pick one of these children at random and call it v_1.

iii. Give the other children independent Galton-Watson descendant trees and give the chosen v_1 an independent number of children $\widehat{X}_2 \sim X$.

iv. Again, pick one of the children of v_1 at random and call it v_2.

v. Repeat this procedure indefinitely to obtain $(\widehat{T}, (v_n)_{n\in\mathbb{N}_0})$. Notice that, since $P(\widehat{X} \geq 1) = 1$, it is indeed possible to repeat this procedure indefinitely. In particular, $\widehat{T}$ is always infinite, i.e. with probability one there is no extinction.

[2]Again, for the sake of this heuristics, we assume that the distribution of X is such that it is indeed possible to have such an urn.

Let us check that the pair $(\widehat{T}, (v_n)_{n\in\mathbb{N}_0})$ constructed like this indeed satisfies (3.1). To this end, fix $n \in \mathbb{N}_0$ and let t be a rooted tree which has at least $n+1$ generations, whose root has exactly k children with corresponding descendant trees $t^{(1)}, \dots, t^{(k)}$. For any vertex $v \in \partial t_{n+1}$, there exists a unique $i_v \in \{1, \dots, k\}$ such that v belongs to the subtree $t^{(i_v)}$. By the construction of T and $\widehat{T}$, it follows that

$$P(T_{n+1} = t_{n+1}) = p_k \prod_{i=1}^{k} P(T_n = t_n^{(i)}) = kp_k \cdot \frac{1}{k} \cdot P(T_n = t_n^{(i_v)}) \cdot \prod_{i \neq i_v} P(T_n = t_n^{(i)})$$

and

$$P(\widehat{T}_{n+1} = t_{n+1} \, , \, v_{n+1} = v) = \frac{kp_k}{\mu} \cdot P(\widehat{T}_n = t_n^{(i_v)} \, , \, v_n = v) \cdot \prod_{i \neq i_v} P(T_n = t_n^{(i)}),$$

so that

$$\frac{P(\widehat{T}_{n+1} = t_{n+1} \, , \, v_{n+1} = v)}{P(T_{n+1} = t_{n+1})} = \frac{1}{\mu} \cdot \frac{P(\widehat{T}_n = t_n^{(i_v)} \, , \, v_n = v)}{P(T_n = t_n^{(i_v)})}. \tag{3.2}$$

Proceeding by induction, we conclude that for any $n \in \mathbb{N}$

$$P(\widehat{T}_n = t_n \, , \, v_n = v) = \frac{1}{\mu^n} P(T_n = t_n).$$

From this we can see the following:

i. By summing over all $v \in \partial t_n$, we get that

$$P(\widehat{T}_n = t_n) = W_n(t) P(T_n = t_n), \tag{3.3}$$

where $W_n(t) := \frac{Z_n(t)}{\mu^n}$. In particular, we have $Z_n(\widehat{T}) \sim \widehat{Z}_n$, with $\widehat{Z}_n$ the size-biasing of Z_n. This is the reason why we call $\widehat{T}$ the "size-biased" tree.

ii. Using (i) we deduce that, conditionally on $\widehat{T}_n$, v_n is chosen uniformly among those in $\partial \widehat{T}_n$. Indeed, for each $v \in \partial t_n$ we have

$$P(v_n = v \,|\, \widehat{T}_n = t_n) = \frac{1}{Z_n(t)}.$$

In the sequel, we will call $\hat{T}$ the size-biased Galton-Watson tree with offspring distribution X.

Remark 3.2.1. *Whenever $\mu \leq 1$ there exists an additional probabilistic interpretation for $\widehat{T}$. Indeed, in this case we have that the distribution of $\widehat{T}$ is that of the original tree T but conditioned to survive forever. Since survival when $\mu \leq 1$ is an event of zero probability, by this we mean that for any fixed $k \in \mathbb{N}$ and rooted tree t one has*

$$P(\widehat{T}_k = t_k) = \lim_{n \to +\infty} P(T_k = t_k | Z_n > 0).$$

We shall prove this fact later in Chapter 6.

3.2.1 A small digression on the distribution of random trees

Recall that if X is a random variable on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$, we define its *distribution* as the probability measure P_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given for each $B \in \mathcal{B}(\mathbb{R})$ by

$$P_X(B) := P(X \in B).$$

We wish to define the analogue of P_X but this time for the random GW-tree T. The way to do this, is to view T as a "random variable" taking values in the space of rooted trees. Indeed, consider:

- $\mathcal{T}$, the space of locally finite rooted trees, i.e. each vertex has finitely many neighbors.
- $\mathcal{T}_n$, the restriction of $\mathcal{T}$ to trees of at most n generations.
- $\mathcal{F}_n^{\mathcal{T}}$, the σ-algebra generated by the events $\{T \in \mathcal{T} : T_n = t\}$ with $t \in \mathcal{T}_n$.
- $\mathcal{F}^{\mathcal{T}} := \sigma(\mathcal{F}_n^{\mathcal{T}} : n \in \mathbb{N})$, the σ-algebra generated by $\cup_{n \in \mathbb{N}} \mathcal{F}_n^{\mathcal{T}}$.

We then define the *distribution* of T as the probability P_T on $(\mathcal{T}, \mathcal{F}^{\mathcal{T}})$ given, for each $B \in \mathcal{F}^{\mathcal{T}}$, by

$$P_T(B) := P(T \in B).$$

It can be seen that P_T is well-defined, in the sense that the value of $P(T \in B)$ does not depend on the particular choice of underlying Galton-Watson process (as long as one only considers processes with the same offspring distribution, of course). The distribution $P_{\widehat{T}}$ of $\widehat{T}$ is defined analogously.

3.3 Branching processes with immigration

The vertices off the distinguished path $(v_0, v_1, \dots)$ of the size-biased tree $\widehat{T}$ form what is known as a *Galton-Watson branching process with immigration* (or GWI for short). In general, such a process is defined via a pair (X, Y) of distributions: the offspring distribution X with which individuals have children, and the immigration distribution Y which describes the arrival of new individuals to the population. The process is defined as follows:

i. It starts with zero particles.

ii. At every generation $n \geq 1$, there is an immigration of Y_n particles, where $(Y_n)_{n \in \mathbb{N}}$ are i.i.d. with distribution Y.

iii. On top of this, each particle in the n-th generation is endowed with an independent ordinary Galton-Watson descendant tree with offspring distribution X.

If for each $n \in \mathbb{N}$ one thinks of those unselected children of the distinguished vertex v_{n-1} from $\widehat{T}$ as the particles immigrating in generation n, one sees that the sequence $(G_n)_{n \in \mathbb{N}}$ defined by

$$G_n := Z_n(\widehat{T}) - 1, \tag{3.4}$$

corresponds to the different generation sizes in a GWI with distribution pair $(X, Y := \widehat{X} - 1)$. Here, X simply denotes the offspring distribution of the original Galton-Watson process.

The advantage of introducing size-biased trees in the study of (regular) GW-processes is that it allows us to establish a link between the regular GW-process and an a particular GWI via (3.4). As we shall see, GWIs are easier to study than regular GWs due to their additional independence. Indeed, in each generation of a GWI there exists a portion of its individuals, those who immigrate exactly in that generation, which is completely independent of what happened in past generations. It is this added independence, which is not present in regular GW-processes, what will make them easier to understand. As proof of this, we have the following result by Seneta which we shall use in the study of the supercritical regime (of regular GW-processes).

Theorem 3.3.1. *Let $(G_n)_{n\in\mathbb{N}}$ be the generation sizes of a GWI with distribution pair (X, Y). Suppose that $\mu := \mathbb{E}(X) \in (1, +\infty)$. Then, we have the following dichotomy:*

- *If $\mathbb{E}(\log^+ Y) < +\infty$ then $\lim_{n\to+\infty} \frac{G_n}{\mu^n}$ exists and is finite almost surely.*
- *If $\mathbb{E}(\log^+ Y) = +\infty$ then $\limsup_{n\to+\infty} \frac{G_n}{c^n} = +\infty$ for any $c > 1$ almost surely.*

3.3.1 Proof of Theorem 3.3.1

We begin the proof by giving an auxiliary lemma.

Lemma 3.3.2. *Let $(X_n)_{n\in\mathbb{N}}$ be nonnegative i.i.d. random variables. Then, almost surely one has*

$$\limsup_{n\to+\infty} \frac{1}{n} X_n = \begin{cases} 0 & \text{if } \mathbb{E}(X_1) < +\infty \\ +\infty & \text{if } \mathbb{E}(X_1) = +\infty. \end{cases}$$

In particular, we have that:

- *If $\mathbb{E}(X_1) < +\infty$ then $\sum_{n\in\mathbb{N}} e^{X_n} c^n < +\infty$ almost surely for all $c \in (0, 1)$.*
- *If $\mathbb{E}(X_1) = +\infty$ then $\limsup_{n\to+\infty} \frac{e^{X_n}}{c^n} = +\infty$ for all $c > 1$ almost surely.*

Proof. The result follows from an application of the Borel-Cantelli lemma, by using the fact that for any random variable X taking values in $\mathbb{N}_0$ one has the formula

$$\mathbb{E}(X) = \sum_{n\in\mathbb{N}} P(X \geq n),$$

which follows from the representation $X = \sum_{n\in\mathbb{N}} \mathbb{1}_{\{X\geq n\}}$ and the monotone convergence theorem. Details are left as an exercise. □

Let us now prove Theorem 3.3.1. Assume first that $\mathbb{E}(\log^+ Y) = +\infty$. Then, by Lemma 3.3.2 we have that $\limsup_{n\to+\infty} \frac{Y_n}{c^n} = +\infty$ for any $c > 1$ almost surely. Since $G_n \geq Y_n$ by construction of the GWI, the result in this case now follows.

Now, let us suppose that $\mathbb{E}(\log^+ Y) < +\infty$ and show that $\lim_{n\to+\infty} W_n$ exists and is finite a.s., where for each $n \in \mathbb{N}$ we write

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

Note that it will be enough to show that

$$P\left((W_n)_{n\in\mathbb{N}} \text{ is a Cauchy sequence}\right) = 1.$$

If $\vec{Y} = (Y_n)_{n\in\mathbb{N}}$ denotes the sequence of immigrating particles, by conditioning on $\vec{Y}$ we get that

$$P\left((W_n)_{n\in\mathbb{N}} \text{ is Cauchy}\right) = \mathbb{E}(g(\vec{Y}))$$

where

$$g(\vec{y}) := P\left((W_n[\vec{y}])_{n\in\mathbb{N}} \text{ is a Cauchy sequence}\right)$$

and $(W_n[\vec{y}])_{n\in\mathbb{N}}$ are the normalized generation sizes of a GWI with **fixed** immigration $\vec{y} = (y_n)_{n\in\mathbb{N}}$. We now wish to show the implication

$$\sum_{n\in\mathbb{N}} \frac{y_n}{\mu^n} < +\infty \Longrightarrow g(\vec{y}) = 1. \tag{3.5}$$

Since by Lemma 3.3.2 we have that $\sum_{n\in\mathbb{N}} \frac{Y_n}{\mu^n} < +\infty$ with probability one, (3.5) then implies that

$$P\left((W_n)_{n\in\mathbb{N}} \text{ is Cauchy}\right) = \mathbb{E}(g(\vec{Y})\mathbb{1}_{\{\sum_{n\in\mathbb{N}} \frac{Y_n}{\mu^n} < +\infty\}}) = P\left(\sum_{n\in\mathbb{N}} \frac{Y_n}{\mu^n} < +\infty\right) = 1,$$

concluding the proof. So, let us show (3.5). By the martingale convergence theorem, it will suffice to show that the sequence $(W_n[\vec{y}])_{n\in\mathbb{N}}$ is a nonnegative submartingale with bounded first moment. Now, given $n \in \mathbb{N}$ observe that, since $y_{n+1} \geq 0$, one has that

$$\begin{aligned}
\mathbb{E}(W_{n+1}[\vec{y}]|W_n[\vec{y}],\dots,W_1[\vec{y}]) &= \mathbb{E}\left(\frac{1}{\mu^{n+1}}\left(\sum_{i=1}^{G_n[\vec{y}]} X_i^{(n)} + y_{n+1}\right)\middle| G_n[\vec{y}],\dots,G_1[\vec{y}]\right) \\
&= \frac{1}{\mu^{n+1}}(\mathbb{E}(X_1^{(n+1)})G_n[\vec{y}] + y_{n+1}) \\
&\geq W_n[\vec{y}],
\end{aligned}$$

which shows that $(W_n[\vec{y}])_{n\in\mathbb{N}}$ is a submartingale. To see that it is also bounded in L^1, for $k < n$ let $G_{n,k}[\vec{y}]$ be the number of descendants in generation n of the y_k particles that immigrated in generation k, so that $G_n[\vec{y}] = \sum_{k=1}^n G_{n,k}[\vec{y}]$ (where we use the convention $G_{n,n}[\vec{y}] := y_n$). Then,

$$\mathbb{E}(W_n[\vec{y}])) = \mathbb{E}\left(\frac{1}{\mu^n}\sum_{k=1}^n G_{n,k}[\vec{y}]\right) = \sum_{k=1}^n \frac{1}{\mu^k}\mathbb{E}\left(\frac{G_{n,k}[\vec{y}]}{\mu^{n-k}}\right).$$

Observe that $G_{n-k}[\vec{y}]$ is just the number of individuals in the $(n-k)$-th generation of an ordinary Galton-Watson process but starting instead with y_k particles. Thus, we have that

$$\mathbb{E}\left(\frac{G_{n,k}[\vec{y}]}{\mu^{n-k}}\right) = y_k,$$

which yields

$$\mathbb{E}\left(W_n[\vec{y}]\right) = \sum_{k=1}^n \frac{y_k}{\mu^k}.$$

In particular, if $\sum_{k\in\mathbb{N}} \frac{y_k}{\mu^k} < +\infty$ we see that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left(W_n[\vec{y}]\right) = \sum_{k\in\mathbb{N}} \frac{y_k}{\mu^k} < +\infty,$$

which shows that $(W_n[\vec{y}])_{n\in\mathbb{N}}$ is bounded in L^1. This concludes the proof.

3.4 Exercises

1. Prove Lemma 3.3.2.

Chapter 4

The supercritical regime

We now turn to the study of the Galton-Watson process in the supercritical regime $\mu \in (1, +\infty)$.

4.1 The Kesten-Stigum theorem

The natural question which appears in this context is what is the rate of growth of Z_n on the event of survival of the branching process. Notice that, since the sequence $(W_n)_{n\in\mathbb{N}}$ with $W_n := \frac{Z_n}{\mu^n}$ is a nonnegative martingale (see Ex. 1 of Section 2.5), it converges almost surely to a finite limit W_∞.

Hence, we see that, whenever $W_\infty > 0$ we get that Z_n has the same order of magnitude as its expected value μ^n (modulo a random factor $W_\infty > 0$), while if $W_\infty = 0$ then it grows more slowly. The question is then the following:

Q1. When exactly is $W_\infty > 0$?

It is clear that on the event of extinction one has that $W_\infty \equiv 0$ but, is it possible to have $W_\infty = 0$ also on the event that the population survives forever? In order to answer this question, we first note the following $0-1$ law for W_∞.

Proposition 4.1.1. *Either $W_\infty \equiv 0$ a.s. or $W_\infty > 0$ a.s. on the event of survival. In other words, we have that $P(W_\infty = 0) \in \{\eta, 1\}$.*

Proof. Notice that for any $n \in \mathbb{N}$ we have the decomposition

$$W_n = \frac{1}{\mu} \sum_{i=1}^{Z_1} W_{n-1}^{(i)} \tag{4.1}$$

where each $W^{(i)} = (W_n^{(i)})_{n\in\mathbb{N}}$ is the Galton-Watson martingale corresponding to the i-th children of the original individual. Taking the limit as $n \to +\infty$ on (4.1) yields

$$W_\infty = \frac{1}{\mu} \sum_{i=1}^{Z_1} W_\infty^{(i)} \tag{4.2}$$

where $\{W_\infty^{(i)} : i = 1, \ldots, Z_1\}$ are i.i.d. random variables given Z_1, distributed according to W_∞. Hence, if we write $\xi := P(W_\infty = 0)$ then, since

$$W_\infty = 0 \Longleftrightarrow W_\infty^{(i)} = 0 \text{ for all } i = 1, \ldots, Z_1,$$

we see that

$$\xi = \mathbb{E}(P(W_\infty = 0|Z_1)) = \mathbb{E}\left(\prod_{i=1}^{Z_1} P(W_\infty^{(i)} = 0)\right) = \mathbb{E}(\xi^{Z_1}) = G_X(\xi),$$

from where it follows that $\xi \in \{\eta, 1\}$ by Theorem 1.3.1. □

In response to (Q1), we have the following result.

Theorem 4.1.2 (Kesten-Stigum(1966))**.** *If $\mu > 1$ then the following statements are equivalent:*

i. $P(W_\infty = 0) = \eta$, *i.e.* $W_\infty > 0$ *a.s. on the event of survival.*

ii. $\mathbb{E}(W_\infty) = 1$, *i.e.* $W_n \xrightarrow{L^1} W_\infty$ *(by Scheffe's lemma).*

iii. $\mathbb{E}(X \log^+ L) < +\infty$.

This result tells us that whenever the offspring distribution is slightly more integrable than just a finite first moment, generations sizes have the same order of magnitude as their means on the event of survival, i.e $Z_n \approx \mu^n = \mathbb{E}(Z_n)$. In other words, if, whenever (iii) holds, we plot $\log Z_n$ as a function of n (and we think of n as a continuous variable for simplicity), on the event of survival we obtain a curve which is eventually almost a straight line with slope $\log \mu$. The intercept of this limiting line on the vertical axis is then the random variable $\log W_\infty$ (which is well-defined here since one has $W_\infty > 0$ on the event of survival). We stress the fact that W_∞ is truly random, even on the event of survival. This randomness arises because the size of the population may undergo proportionally large random fluctuations in the early generations, and these fluctuations act as multiplicative factors whose effect persists in later generations. Indeed, one way to see this is by noticing that for any $k \in \mathbb{N}$ we have

$$\mathbb{E}(W_\infty|Z_1 = k) = \frac{1}{\mu}\sum_{i=1}^{k} \mathbb{E}(W_\infty^{(i)}|Z_1 = k) = \frac{k}{\mu}\mathbb{E}(W_\infty) = \frac{k}{\mu}$$

by virtue of the representation in (4.2), which shows that W_∞ is sensible to the value of Z_1 (and, similarly, also to the value of all early generations).

However, whenever (iii) in Theorem 4.1.2 fails, the mean μ^n overestimates the rate of growth. The following question then arises:

Q2. What is the rate of growth of Z_n whenever the population survives but $W_\infty = 0$?

The following result shows that, whenever (iii) fails, there is still an essentially deterministic rate of growth, as shown by Seneta (1968) and Heyde (1970), which is only slightly less than μ^n.

Theorem 4.1.3 (Seneta-Heyde)**.** *If $\mu > 1$ then there exist constants $(c_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}_{>0}$ such that:*

i. $\widetilde{W}_\infty := \lim_{n\to+\infty} \frac{Z_n}{c_n}$ *exists and is finite almost surely.*

ii. $P(\widetilde{W}_\infty = 0) = \eta$, *i.e.* $\widetilde{W}_\infty > 0$ *a.s. on the event of survival.*

iii. $\lim_{n\to+\infty} \frac{c_{n+1}}{c_n} = \mu$.

Proof. See [4, Chapter 5]. □

4.2 Proof of Theorem 4.1.2

The proof of the Kesten-Stigum theorem is a combination of a general decomposition lemma from measure theory together with the results on GWIs previously established in Chapter 3.

4.2.1 The decomposition lemma

Lemma 4.2.1. *Let μ be a finite measure and ν be a probability on some measurable space $(\Omega, \mathcal{F})$. Suppose that $(\mathcal{F}_n)_{n\in\mathbb{N}}$ are an increasing sequence of sub-σ-algebras whose union generates $\mathcal{F}$ and that for each $n \in \mathbb{N}$ we have $\mu|_{\mathcal{F}_n} \ll \nu|_{\mathcal{F}_n}$ with $X_n := \frac{d\mu|_{\mathcal{F}_n}}{d\nu|_{\mathcal{F}_n}}$. Finally, define $X := \limsup_{n\to+\infty} X_n$. Then,*

$$\mu \ll \nu \iff X < +\infty \;\; \mu\text{-a.s.} \iff \int_\Omega X d\nu = \mu(\Omega)$$

and

$$\mu \perp \nu \iff X = +\infty \;\; \mu\text{-a.s.} \iff \int_\Omega X d\nu = 0.$$

Proof. First, notice that the sequence $(X_n)_{n\in\mathbb{N}}$ is a ν-martingale. Indeed, by definition of X_n and the fact that μ is finite, we have that

$$\mathbb{E}(X_n) = \int_\Omega X_n d\nu = \mu(\Omega) < +\infty$$

which shows condition (M1) in the definition of martingale. Now, to check (M2) we observe that, since each X_k is $\mathcal{F}_k$-measurable by construction and the sequence $(\mathcal{F}_k)_{k\in\mathbb{N}}$ is increasing, one has the inclusion $\sigma(X_1, \ldots, X_n) \subseteq \mathcal{F}_n$ for all $n \in \mathbb{N}$ and thus, given $A \in \sigma(X_1, \ldots, X_n)$, we see that

$$\int_A X_{n+1} d\nu = \mu(A) = \int_A X_n d\nu,$$

by definition of X_{n+1} and X_n, since $A \in \mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. This shows (M2).

In particular, being also nonnegative, by the martingale convergence theorem we know that $(X_n)_{n\in\mathbb{N}}$ converges ν-a.s. to $X := \limsup_{n\to+\infty} X_n$ and that X is ν-a.s. finite. We now claim that the measure μ has the following decomposition: for any $A \in \mathcal{F}$

$$\mu(A) = \int_A X d\nu + \mu(A \cap \{X = +\infty\}). \tag{4.3}$$

Equivalently, with the notation from (2.3), our claim is that

$$\mu = X \cdot \nu + \mathbb{1}_{\{X=+\infty\}} \cdot \mu.$$

Given this, the lemma immediately follows. Indeed:

- If $\mu \ll \nu$ then $X < +\infty$ μ-a.s.; if $X < +\infty$ μ-a.s. then by (4.3) we see that $\mu(\Omega) = \int_\Omega X d\nu$; and if $\int_\Omega X d\nu = \mu(\Omega)$ then $X < +\infty$ μ-a.s. by (4.3) and $\mu = X \cdot \nu$, so that $\mu \ll \nu$.
- If $\mu \perp \nu$ then by (4.3) we have $\mu = \mathbb{1}_{\{X=+\infty\}} \cdot \mu$ and so $X = +\infty$ μ-a.s.; if $X = +\infty$ μ-a.s. then $X \cdot \nu = 0$, and so $\int_\Omega X d\nu = 0$; and, finally, if $\int_\Omega X d\nu = 0$ then $X = +\infty$ μ-a.s. by (4.3), so that $\mu \perp \nu$.

To show (4.3) suppose first that $\mu \ll \nu$. Then, $X < +\infty$ μ-a.s. so that $\mathbb{1}_{\{X=+\infty\}} \cdot \mu = 0$. Moreover, if $\bar{X} := \frac{d\mu}{d\nu}$ then by definition of X_n we have $X_n = \mathbb{E}(\bar{X}|\mathcal{F}_n)$ for each n. Hence, by Theorem 2.2.2 it follows that $(X_n)_{n\in\mathbb{N}}$ converges ν-a.s. to $\bar{X}$. In particular, we have that $X = \bar{X}$ ν-almost surely. But, since $\mu = \bar{X} \cdot \nu$ by definition of $\bar{X}$, it follows from this that

$$\mu = X \cdot \nu = X \cdot \nu + \mathbb{1}_{\{X=+\infty\}} \cdot \mu$$

which gives the desired decomposition.

Now, to treat the general case, we define the probability measure

$$\rho := \frac{1}{C}(\mu + \nu)$$

with $C := (\mu + \nu)(\Omega) > 0$ and observe that $\mu, \nu \ll \rho$, so that we may apply what we have already shown to the variables $U_n := \frac{d\mu|_{\mathcal{F}_n}}{d\rho|_{\mathcal{F}_n}}$ and $V_n := \frac{\nu|_{\mathcal{F}_n}}{d\rho|_{\mathcal{F}_n}}$. Indeed, define $U := \limsup_{n\to+\infty} U_n$ and $V := \limsup_{n\to+\infty} V_n$. By what we have already shown, we know that ρ-a.s. $U_n \to U$ and $V_n \to V$. On the other hand, since one can easily verify that both $U_n + V_n$ and C are versions of $\mathbb{E}_\rho(C|\mathcal{F}_n)$, we have that ρ-a.s. $U_n + V_n = C$, which itself implies that $\rho(U = V = 0)$. Thus, we see that ρ-a.s.

$$\frac{U}{V} = \frac{\lim_{n\to+\infty} U_n}{\lim_{n\to+\infty} V_n} = \lim_{n\to+\infty} \frac{U_n}{V_n} = \lim_{n\to+\infty} \frac{X_n V_n}{V_n} = \lim_{n\to+\infty} X_n = X,$$

where we have used that ρ-a.s. $U_n = X_n V_n$, which follows from the definitions of U_n, V_n and X_n. Therefore, using three times what we established before, we obtain

$$\mu = U \cdot \rho = XV \cdot \rho + (\mathbb{1}_{\{V=0\}} U) \cdot \rho = X \cdot \nu + \mathbb{1}_{\{X=+\infty\}} \cdot \mu,$$

which concludes the proof. □

4.2.2 Conclusion of the proof

Let P_T and $P_{\widehat{T}}$ denote the distributions of T and $\widehat{T}$, respectively. Then, following the notation of Section 3.2.1, (3.3) can be restated as

$$\frac{dP_{\widehat{T}}|_{\mathcal{F}_n^{\mathcal{T}}}}{dP_T|_{\mathcal{F}_n^{\mathcal{T}}}}(t) = W_n(t)$$

for each $t \in \mathcal{T}$ and $n \in \mathbb{N}$. Now, define for $t \in \mathcal{T}$

$$W(t) := \limsup_{n\to+\infty} W_n(t).$$

By Lemma 4.2.1, we have that

$$\int_{\mathcal{T}} W dP_T = 1 \iff W < +\infty \;\; P_{\widehat{T}}\text{-a.s.} \iff P_{\widehat{T}} \ll P_T \tag{4.4}$$

and

$$W = 0 \;\; P_T\text{-a.s.} \iff W = +\infty \;\; P_{\widehat{T}}\text{-a.s.} \iff P_{\widehat{T}} \perp P_{T^Z}. \tag{4.5}$$

This is dichotomy is crucial because it allows us to change the problem from one about the behavior of the random variable W under P_T to one about its behavior under $P_{\widehat{T}}$, which is easier to handle. Indeed, since the $P_{\widehat{T}}$-behavior of W is described by Theorem 3.3.1, the Kesten-Stigum theorem is now immediate:

- If $\mathbb{E}(X \log^+ X) < +\infty$ then $\mathbb{E}(\log^+ \widehat{X}) < +\infty$, so that $W < +\infty$ $P_{\mathcal{T}z}$-a.s. by Theorem 3.3.1. By (4.4), this implies that $\int_{\mathcal{T}} W dP_{\mathcal{T}} = 1$. But, since the $P_{\mathcal{T}}$-distribution of W is the same as that of W_∞ in the statement of the theorem[1], this means that $\mathbb{E}(W_\infty) = 1$. In particular, W_∞ cannot be zero P-a.s., and so by Proposition 4.1.1 we have that $P(W_\infty = 0) = \eta$.

- If $\mathbb{E}(X \log^+ X) = +\infty$ then $\mathbb{E}(\log^+ \widehat{X}) = +\infty$, so that $W = +\infty$ $P_{\mathcal{T}z}$-a.s. by Theorem 3.3.1. By (4.5), this implies that $W = 0$ $P_{\mathcal{T}z}$-a.s., so that $\mathbb{E}(W_\infty) = 0$ and $P(W_\infty = 0) = 1$.

[1]In general, any measurable function $f : \mathcal{T} \to \mathbb{R}$ under $P_{\mathcal{T}}$ has the same distribution as $f(T)$ under P.

4.3 Exercises

1. (Scheffe's Lemma) Suppose that $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables with $X_n \xrightarrow{as} X$. Show that if $\mathbb{E}(|X|) < +\infty$ then

$$X_n \xrightarrow{L^1} X \iff \mathbb{E}(|X_n|) \longrightarrow \mathbb{E}(|X|).$$

Chapter 5

The subcritical regime

When a Galton-Watson process is subcritical/critical, asking about its asymptotic rate of growth as we did before does not make much sense since the population eventually dies out almost surely. However, there are new questions that one can now ask, such as: how quickly does the population die out? One way to quantify this is by giving the decay rate of $P(Z_n > 0)$. Indeed, notice that, if τ denotes the extinction time of the population given by

$$\tau := \inf\{n \in \mathbb{N} : Z_n = 0\},$$

then

$$\{Z_n > 0\} = \{\tau > n\},$$

so that any information on the decay rate of $P(Z_n > 0)$ immediately translates into information on the (distributional) tail of τ.

A simple estimate for $P(Z_n > 0)$ in the subcritical case is given by Markov's inequality:

$$P(Z_n > 0) = P(Z_n \geq 1) \leq \mathbb{E}(Z_n) = \mu^n.$$

Notice that this estimate already yields the following bound on the expected extinction time:

$$\mathbb{E}(\tau) = \sum_{n=0}^{\infty} P(\tau > n) \leq \sum_{n=0}^{\infty} \mu^n = \frac{1}{1-\mu}.$$

The question is then the following:

Q3. When is μ^n the correct decay rate for $P(Z_n > 0)$?

In response to (Q3), we have the following result.

Theorem 5.0.1 (Heathcote, Seneta and Vere-Jones (1967)). *The sequence $(a_n)_{n\in\mathbb{N}}$ given by*

$$a_n := \frac{P(Z_n > 0)}{\mu^n}$$

is decreasing for any Galton-Watson process with $\mu \in (0, +\infty)$. If, in addition, one has that $\mu < 1$, then the following statements are equivalent:

i. $\lim_{n\to+\infty} a_n > 0$,

ii. $\sup_{n\in\mathbb{N}} \mathbb{E}(Z_n | Z_n > 0) < +\infty$,

iii. $\mathbb{E}(X \log^+ X) < +\infty$.

5.1 Proof of Theorem 5.0.1

We will use a similar approach to that of the previous chapter: we combine two general lemmas together with a result for BPIs.

5.1.1 Two auxiliary lemmas

We now give the two auxiliary lemmas we shall need for the proof. Their proofs are left as exercises.

Lemma 5.1.1. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables taking values on $\mathbb{N}_0$, each of them with a finite mean $x_n := \mathbb{E}(X_n)$. Then,*

- $(\widehat{X}_n)_{n\in\mathbb{N}}$ *tight* $\Longrightarrow \sup_{n\in\mathbb{N}} x_n < +\infty$.
- $\widehat{X}_n \xrightarrow{P} +\infty \Longrightarrow \lim_{n\to+\infty} x_n = +\infty$.

Lemma 5.1.2. *Let $(Y_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random variables. Then,*

$$P\left(\sum_{n\in\mathbb{N}} Y_n = +\infty\right) \in \{0,1\}.$$

5.1.2 An auxiliary result on GWIs

The proof of Theorem 5.0.1 will rely on the following result on GWIs due to Heathcote.

Theorem 5.1.3. *Let $(G_n)_{n\in\mathbb{N}}$ be the generation sizes in a GWI having distribution pair (X,Y). If $\mu = \mathbb{E}(X) < 1$ then we have the following dichotomy:*

- *If $\mathbb{E}(\log^+ Y) < +\infty$ then $(G_n)_{n\in\mathbb{N}}$ converges in distribution to a finite random variable G.*
- *If $\mathbb{E}(\log^+ Y) = +\infty$ then $(G_n)_{n\in\mathbb{N}}$ converges in probability to $+\infty$.*

Proof. For any $k \leq n \in \mathbb{N}$, consider again $G_{n,k}$, the number of descendants of generation n coming from the Y_k individuals that immigrated in generation k, so that $G_n = \sum_{k=1}^n G_{n,k}$. Observe that, since the random variables $G_{n,k}$ are independent for different values of k and their distributions only depend on $n-k$, we have $G_n \sim \sum_{k=1}^n G_{2k-1,k}$. The latter sum increases in n to

$$G := \sum_{k\in\mathbb{N}} G_{2k-1,k}.$$

By Lemma 5.1.2, G is either a.s. finite or a.s. infinite. Thus, to prove the result it suffices to show that G is a.s. finite if and only if $\mathbb{E}(\log^+ Y) < +\infty$.

Now, assume first that $\mathbb{E}(\log^+ Y) < +\infty$. If $\vec{Y} = (Y_k)_{k\in\mathbb{N}}$ is the vector of immigrating particles, then by the (conditional) monotone convergence theorem,

$$\mathbb{E}(G|\vec{Y}) = \sum_{k\in\mathbb{N}} Y_k \mu^k. \tag{5.1}$$

By Lemma 3.3.2, we know that the right-hand side of (5.1) is almost surely finite, so that $\mathbb{E}(G|\vec{Y})$ is also almost surely finite. Since by linearity of the conditional expectation we have that

$$\mathbb{E}(G|\vec{Y}) = \mathbb{E}(G\mathbb{1}_{G<+\infty}|\vec{Y}) + \infty \cdot P(G = +\infty|\vec{Y}),$$

the a.s. finiteness of $\mathbb{E}(G|\vec{Y})$ implies that almost surely

$$P(G = +\infty|\vec{Y}) = 0.$$

Taking expectations on this equality shows that G is almost surely finite.

To check the other implication, it will suffice to show that if G is almost surely finite then

$$\sum_{k\in\mathbb{N}} Y_k(P(X \geq 1))^{k-1} < +\infty \;\; P\text{-almost surely.} \tag{5.2}$$

By Lemma 3.3.2, this will then imply that $\mathbb{E}(\log^+ Y) < +\infty$.

To show (5.2) whenever G is almost surely finite, we first observe that we can write

$$G = \sum_{k\in\mathbb{N}} \sum_{i=1}^{Y_k} Z^{(k)}_{k-1}(i)$$

where each $Z^{(k)}_{k-1}(i)$ is the size of generation $k-1$ in the Galton-Watson subtree corresponding to the i-th of the Y_k immigrating individual in generation k. In particular, if G is a.s. finite then

$$1 = P(G < +\infty) = \mathbb{E}(P(G < +\infty|\vec{Y})) = \mathbb{E}(h(\vec{Y})) \tag{5.3}$$

where, for each $\vec{y} := (y_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}_0$, we define

$$h(\vec{y}) := P(G[\vec{y}] < +\infty)$$

with $G[\vec{y}]$ given by

$$G[\vec{y}] := \sum_{k\in\mathbb{N}} \sum_{i=1}^{y_k} Z^{(k)}_{k-1}(i)$$

for $\{Z^{(k)}_{k-1}(i) : k \in \mathbb{N}, i = 1, \dots, y_k\}$ an array of independent random variables with each $Z^{(k)}_{k-1}(i)$ distributed according to Z_{k-1}, the number of individuals in generation $k-1$ of a GW-process with offspring distribution X. Since $h \leq 1$ by definition, it follows from (5.3) that

$$P(h(\vec{Y}) = 1) = 1. \tag{5.4}$$

Now, let us suppose that $\vec{y} = (y_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}_0$ is such that $G[\vec{y}]$ is almost surely finite, i.e. $h(\vec{y}) = 1$. Since each $Z^{(k)}_{k-1}(i)$ is integer valued, this implies that, with probability one, only finitely many of the random variables $\{Z^{(k)}_{k-1}(i) : k \in \mathbb{N}, i = 1, \dots, y_k\}$ are nonzero. By virtue of their independence, the Borel-Cantelli lemma implies that

$$\sum_{k\in\mathbb{N}} \sum_{i=1}^{y_k} P(Z^{(k)}_{k-1}(i) \neq 0) < +\infty.$$

Since $P(Z^{(k)}_{k-1}(i) \neq 0) \geq (P(X \geq 1))^{k-1}$, we conclude that

$$\sum_{k\in\mathbb{N}} y_k(P(X \geq 1))^{k-1} < +\infty.$$

In particular, what we have shown implies that

$$h(\vec{Y}) = 1 \Longrightarrow \sum_{k\in\mathbb{N}} Y_k(P(X \geq 1))^{k-1} < +\infty. \tag{5.5}$$

Since the left-hand side in (5.5) occurs almost surely by (5.4), this implies that (5.2) holds and thus concludes the proof. □

5.1.3 Conclusion of the proof

Let us finish the proof of Theorem 5.0.1. We follow the proof by Lyons, Pemantle and Peres [3]. For any tree t with $Z_1(t) > 0$, let us label the subtrees corresponding to the children of the root as $(t^{(i)} : i = 1, \dots, Z_1(t))$ and define $\xi_n(t)$ as the lowest-labeled of these children which has at least one descendant in generation n, i.e.

$$\xi_n(t) := \min\{i = 1, \dots, Z_1(t) : Z_{n-1}(t^{(i)}) > 0\},$$

with the convention that $\min \emptyset = +\infty$. Further, set $H_n(t)$ as the number of descendants of $\xi_n(t)$ in generation n, with the convention that $t^{(+\infty)}$ is the empty tree, so that $H_n(t) = 0$ if $\xi_n(t) = +\infty$. Now, observe that for any $k \in \mathbb{N}$

$$P(H_n = k | Z_n > 0) = P(Z_{n-1}^{(\xi_n)} = k | Z_1 > 0\,,\, Z_{n-1}^{(\xi_n)} > 0) = P(Z_{n-1} = k | Z_{n-1} > 0) \tag{5.6}$$

Indeed, this follows from the fact that, given $Z_1 > 0$, the subtree corresponding to ξ_n is a GW-tree with offspring distribution X. But, since $H_n \leq Z_n$, (5.6) implies that

$$P(Z_{n-1} \geq k | Z_{n-1} > 0) = P(H_n \geq k | Z_n > 0) \leq P(Z_n \geq k | Z_n > 0) \tag{5.7}$$

for all $k \in \mathbb{N}$. Summing (5.7) over k yields

$$\mathbb{E}(Z_{n-1} | Z_{n-1} > 0) \leq \mathbb{E}(Z_n | Z_n > 0). \tag{5.8}$$

for each $n \in \mathbb{N}$. Finally, by noticing that

$$P(Z_n > 0) = \frac{\mathbb{E}(Z_n)}{\mathbb{E}(Z_n | Z_n)} = \frac{\mu^n}{\mathbb{E}(Z_n | Z_n > 0)}$$

we deduce from (5.8) that the sequence $(a_n)_{n \in \mathbb{N}}$ is monotone decreasing and that (i)$\Longleftrightarrow$(ii).

To check the remaining equivalence, for each $n \in \mathbb{N}$ let ν_n be the distribution of Z_n conditioned on $Z_n > 0$, i.e. for each $k \in \mathbb{N}$

$$\nu_n(k) = \frac{P(Z_n = k)}{P(Z_n > 0)}.$$

Notice that $\widehat{\nu}_n$, the size-biasing of ν_n, coincides with the distribution of $\widehat{Z}_n$. Indeed, for any $k \in \mathbb{N}$ we have that

$$\widehat{\nu}_n(k) = \frac{k \nu_n(k)}{\mathbb{E}(Z_n | Z_n > 0)} = \frac{k P(Z_n = k)}{\mathbb{E}(Z_n)} = P(\widehat{Z}_n = k).$$

Thus, by construction of $\widehat{T}$ we know that $\widehat{\nu}_n \sim G_n + 1$ where $(G_n)_{n \in \mathbb{N}}$ are the generation sizes of a GWI with distribution pair $(X, \widehat{X} - 1)$.

Now, suppose that $\mu < 1$. If (ii) holds, that is, the means of ν_n are uniformly bounded in n, then, by Lemma 5.1.1, $(\widehat{Z}_n)_{n \in \mathbb{N}}$ cannot converge in probability to $+\infty$ and so neither can $(G_n)_{n \in \mathbb{N}}$. Theorem 5.1.3 then implies that $\mathbb{E}(\log^+(\widehat{X} - 1)) < +\infty$, so that (iii) holds. Conversely, if (iii) holds then $(\widehat{X}_n)_{n \in \mathbb{N}}$ is tight by Theorem 5.1.3, so that (ii) then holds by Lemma 5.1.1.

5.2 Exercises

1. Prove Lemma 5.1.1.

2. Prove Lemma 5.1.2.

Bibliography

[1] Jochen Geiger. Elementary new proofs of classical limit theorems for Galton-Watson processes. *J. Appl. Probab.*, 36(2):301–309, 1999.

[2] Theodore E. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.

[3] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.

[4] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016.