Branching Processes, Martingales, Kingman's Subadditive Ergodic Theorem, and some applications

Sascha Troscheit

Department of Pure Mathematics, University of Waterloo, Waterloo, Ont., N2L 3G1, Canada
`stroscheit@uwaterloo.ca`

February 19, 2020

Abstract

These are the lecture notes for a short learning seminar given at the Department of Pure Mathematics, University of Waterloo.

Branching processes are stochastic processes often used to model reproduction and were first developed to study effects in populations such as extinction of surnames and Brownian motion. However, these models have applications across a whole range of pure mathematics and have been employed in solving problems in analysis, combinatorics, number theory, and group theory. In this series of seminars we take a probabilistic and dynamical view on several types of branching processes and study basic properties such as the change of number of descendants. To do this we will use two tools, the theory of martingales and ergodic theory, each of which become applicable in different situations:

- Classical branching process are often martingales and we will state and apply the martingale convergence theorems to define several geometrically interesting measures with applications to models such as percolation.
- For some modern branching processes, we state and prove a version of Kingman's subadditive ergodic theorem. This theorem is a very powerful tool in probability and dynamical systems with many applications to other fields such as number theory. We will apply this theorem to those branching processes to obtain some interesting, albeit implicit, results and discuss some open problems for these processes.

Contents

1 Classical Branching processes

Branching processes are stochastic processes that capture a 'population' where each 'individual' has a given distribution of producing 'descendants'. In effect, they are random trees, where each node has a random number of descendants. The archetypal of such branching process is called the *Galton–Watson process* and assumes that the offspring of an individual is independent and identical in distribution for every individual. We will analyse Galton–Watson processes in this section and continue exploring other branching processes in later sections, following an approach similar to Athreya [AN72]. To start, we recall the definition of a tree and its boundary[1].

Definition 1.1. *Let Λ be a countable index set and let $\Lambda^* = \bigcup_{i=1}^{\infty} \Lambda^i$ be the set of finite words over Λ. A* tree *τ (over Λ) is a subset of Λ^* such that:*

- *For every $v \in \tau$ there exists $\lambda \in \Lambda$ such that $v\lambda \in \tau$.*
- *For every $v \in \tau$ of length greater than one, there exist $\lambda \in \Lambda$ and $w \in \tau$ such that $v = w\lambda$.*

We note that some authors prefer rooted trees, in which case one would require the empty word to be in τ, acting as the root of the entire tree. Contrary to those authors we will allow "empty trees", *i.e.* $\tau = \varnothing$ and random forests.

Definition 1.2. *The* (Gromov) boundary *$\partial\tau$ of a tree τ are all infinite words $v \in \Lambda^{\mathbb{N}}$ such that every finite restriction is in τ,* i.e. *$v|_k = v_1 v_2 \dots v_k \in \tau$ for all $k \in \mathbb{N}$.*

1.1 Galton–Watson Process

We capture the distribution of offspring that an individual produces as a probability vector, called the *offspring distribution.*

Definition 1.3. *The* offspring distribution *is the probability vector $\vec{\theta} = (\theta_0, \theta_1, \dots)$, where $\sum_{i=0}^{\infty} \theta_i = 1$. If additionally $\theta_0 + \theta_1 < 1$, we refer to $\vec{\theta}$ as a* non-trivial offspring distribution*.*

Let X_i^j be the random variable such that $\mathbb{P}\{X_i^j = k\} = \theta_k$. In particular, X_i^j are pairwise independent and have the same distribution. We write X for a generic copy of that random variable.

Definition 1.4. *The sequence of random variables $(Z_0, Z_1, \dots)$ is called a* Galton–Watson process *if $Z_0 = X_0^0$ and*

$$Z_{i+1} = \sum_{j=1}^{Z_i} X_{i+1}^j.$$

This model was originally proposed by to study the extinction of surnames by Sir Francis Galton. He posed his question in the *Educational Times* in 1873, to which Rev. Henry Watson replied. Later they published a joint article on the extinction of families [GW75]. Clearly, if there exists k such that $Z_k = 0$ we must have $Z_l = 0$ for all $l \geq k$ and this extinction is what originally motivated Galton. Similarly, the first question we shall attempt to answer is:

[1] So far, we are not using this definition and it may get deleted or moved later.

Question 1.5. *Given an offspring distribution $\vec{\theta}$, what is the probability that the associated Galton–Watson process becomes extinct? That is, we aim to determine*

$$\mathbb{P}\{Z_k = 0 \text{ for some } k\}.$$

Alternatively, we can define the k-th random variable in the Galton–Watson process as the number of words of length k in a random tree where each node $v \in \tau$ has offspring distribution according to τ, that is, for all $v \in \tau$,

$$\theta_k = \mathbb{P}\{\#\{\lambda \in \Lambda \mid v\lambda \in \tau\} = k\}.$$

By our definition of trees, there is a slight issue comparing offspring in this way: no node can have zero children and τ only contains nodes with children. We shall gloss over this fact now and put it on a formal footing later.

To determine when a Galton–Watson process 'dies out' we will use a powerful tool called the *probability generating function.*

1.2 Probability Generating Function

The probability generating function of a discrete random variable is a power series representation that 'encodes' many useful properties of the random variable. It is defined in terms of expectations, and we write $\mathbb{E}(X) = \int_\Omega X(\omega) d\,\mathbb{P}(\omega)$ for the expectation of X with respect to some probability space $(\Omega, \mathbb{P})$. We will ignore the issue of measurability entirely as all the random variables we will consider are (Borel) measurable.

Definition 1.6. *The* probability generating function $f(X, s)$ *of a discrete random variable* X *is defined as*

$$f(X, s) = \mathbb{E}(s^X) = \sum_{i=0}^{\infty} \mathbb{P}\{X = i\} s^i.$$

For simplicity, we will often write $f(s)$ if $X = Z_0$ is the random variable associated with a given offspring distribution for a Galton–Watson process.

Example. Let $\vec{\theta} = \{p, 1-p, 0, 0, \dots\}$ for some $0 \leq p \leq 1$. If $p = 1$, then clearly the Galton–Watson tree τ is empty almost surely. Similarly, if $p = 0$, then $Z_k = 1$ for all k almost surely. If $0 < p < 1$, then $Z_k = 1$ with probability $(1-p)^{k+1}$ and hence the process dies out almost surely. Since this behaviour is so simple and uninformative we refer to the case $\theta_0 + \theta_1 = 1$ as a trivial Galton–Watson process. In both of these cases the probability generating function is linear, $f(s) = p + (1-p)s$, with $f(0) = \theta_0$ and $f(1) = 1$.

As it turns out, the last two facts hold in much greater generality, but the generating function is now strictly convex, a fact that we will use later on. We write $f^{(k)}$ for the k-th derivative of f and $f_k(s)$ for the k-fold composition of the map f, *i.e.* $f_1(s) = f(s)$, $f_2(s) = (f \circ f)(s)$, $f_3(s) = (f \circ f \circ f)(s)$, *et cætera.* Combining composition and differentiation, we write $f_k^{(l)}(s) = (d^l/ds^l) f_k(s)$.

Theorem 1.7. *Let $\vec{\theta}$ be a non-trivial offspring distribution with associated random variable X. The following statements hold for its generating function $f(s)$.*

1. $f(0) = \theta_0$ *and* $\lim_{s \nearrow 1} f(s) = 1$.
2. $f(s)$ *is smooth on* $[0, 1)$ *and* $f^{(k)}(s)$ *is continuous at* $s = 1$ *if* $f^{(k)}(1) < \infty$.

3. *The offspring distribution can be recovered by*

$$\mathbb{P}\{X = k\} = \frac{f^{(k)}(0)}{k!}.$$

4. *The mean of X is $m = \mathbb{E}(X) = f^{(1)}(1)$.*

5. *Assume $\mathbb{E}(X)$ is finite. Then the variance $\mathrm{Var}(X) = \mathbb{E}((X-m)^2)$ is*

$$\mathrm{Var}(X) = f^{(2)}(1) + m - m^2.$$

6. *$f(s)$ is strictly convex and increasing in s on $[0,1]$*

7. *If $m \leq 1$, then $f(t) > t$ for all $t \in [0,1)$. If $m > 1$, then there exists a unique $q \in [0,1)$ such that $f(q) = q$.*

8. *If $t \in [0,q)$, then $f_k(t) \nearrow q$ as $k \to \infty$. If $t \in (q,1)$, then $f_k(t) \searrow q$ as $k \to \infty$.*

Proof.

1. The first claim is trivial and the second follows from Abel's theorem.

2. Smoothness arises from the non-negativity of entries in the sum (and Abel's theorem for $s = 1$).

3. Simple computation. Differentiating once we get

$$f^{(1)}(s) = \sum_{i=1}^{\infty} \theta_i \, i \, s^{i-1} = \sum_{i=0}^{\infty} \theta_i \, i \, s^{i-1},$$

differentiating again,

$$f^{(2)}(s) = \sum_{i=0}^{\infty} \theta_i \, i \, (i-1) \, s^{i-2},$$

and in general

$$f^{(k)}(s) = \sum_{i=0}^{\infty} \theta_i \, i \, (i-1) \, \ldots \, (i-k+1) s^{i-k}.$$

Setting $s = 0$ gives the desired result.

4. Simple computation $f^{(1)}(1) = \sum_{i=1}^{\infty} \theta_i \, i \, 1^i = \mathbb{E}(X)$.

5. We compute

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}((X-m)^2) &= \sum_{i=0}^{\infty} \theta_i (i-m)^2 = \sum_{i=0}^{\infty} \theta_i \, i^2 + \sum_{i=0}^{\infty} \theta_i \, m^2 - 2 \sum_{i=0}^{\infty} \theta_i \, m \, i \\ &= \sum_{i=0}^{\infty} \theta_i \, i^2 + (-m+m) - m^2 = \sum_{i=0}^{\infty} \theta_i \, i^2 - \sum_{i=0}^{\infty} \theta_i \, i + m - m^2 \\ &= \sum_{i=0}^{\infty} \theta_i \, i(i-1) + m - m^2 = f^{(2)}(1) + m - m^2.\end{aligned}$$

6. Positive constants and non-triviality imply $f^{(1)}(s) > 0$ and $f^{(2)}(s) > 0$, thus $f(s)$ is strictly increasing and convex.

7. Follows from 6.

8. Clearly $0 \leq t < q$ gives $t < f(t) < f(q)$ and iterating gives

$$t < f(t) < f_2(t) < f_3(t) < \cdots < f_k(t) < f(q) = q.$$

By continuity $f_k(t) \nearrow L$ for some L that satisfies $f(L) = L$. But q is the least root and thus $L = q$. The other limit follows similarly. □

1.2.1 Linking the generating function to Z_k

There is a strong connection between iterates of the probability generating function and the Galton–Wason process Z_k. Consider

$$\begin{aligned} f(Z_1, s) &= f\Big(\sum_{i=1}^{Z_0} X_1^i, s\Big) = \mathbb{E}(s^{\sum_{i=1}^{Z_0} X_1^i}) \\ &= \mathbb{E}(\mathbb{E}(s^{X_1^1} s^{X_1^2} \ldots s^{X_1^{Z_0}} \mid Z_0)) \\ &= \mathbb{E}((\mathbb{E}(s^{X_1^1}))^{Z_0}) = \mathbb{E}(f(X_1^1, s)^{X_0^1}) \\ &= f(X, f(X, s)) = f_2(X, s). \end{aligned}$$

This can be extended (in the obvious way) for all $k \in \mathbb{N}$ and $f(Z_k, s) = f_k(X, s)$.

Theorem 1.8. *Let $\{Z_k\}$ be a Galton–Watson process with mean $m = \mathbb{E}(Z_0) = \mathbb{E}(X) < \infty$. Then $\mathbb{E}(Z_k) = m^k$.*

Proof. From Theorem 1.7 and the chain rule we deduce,

$$\begin{aligned} f_1^{(1)}(Z_k, s) = f_k^{(1)}(X, s) &= f_{k-1}^{(1)}(X, f_1(X, s)) \cdot f_1^{(1)}(X, s) \\ &= f_{k-2}^{(1)}(X, f_2(X, s)) \cdot f_1^{(1)}(X, f_1(X, s)) \cdot f_1^{(1)}(X, s) \\ &= f^{(1)}(X, f_{k-1}(X, s)) \cdot f^{(1)}(X, f_{k-2}(X, s)) \cdot \ldots \cdot f^{(1)}(X, s). \end{aligned}$$

But then,

$$\mathbb{E}(Z_k) = f_1^{(1)}(Z_k, 1) = f^{(1)}(X, f_{k-1}(X, 1)) \cdot \ldots \cdot f^{(1)}(X, 1) = [f^{(1)}(X, 1)]^k = m^k. \qquad \square$$

Theorem 1.9. *Let $\{Z_k\}$ be a Galton–Watson process and let $q \in [0, 1]$ be the smallest solution to $f(Z_0, q) = q$. Then the probability of extinction of the process is q,*

$$\mathbb{P}\{Z_k = 0 \mid \textit{ for some } k\} = q$$

Proof. First note that

$$\mathbb{P}\{Z_k = 0 \mid \text{ for some } k\} = \lim_k \mathbb{P}\{Z_k = 0\} = \lim_k f(Z_k, 0)$$

and so by Lemma 1.7,

$$\mathbb{P}\{Z_k = 0 \mid \text{ for some } k\} = \lim_k f_k(X, 0) = q. \qquad \square$$

1.2.2 Example: Mandelbrot percolation

The n-fold Mandelbrot percolation of the d-dimensional unit cube for threshold value $0 < p < 1$ is defined recursively in the following way: Let Q_1 be the set containing the unit cube. The set Q'_{n+1} is defined as the set of all cubes that are obtained by splitting all cubes in Q_n into n^d smaller cubes of the same dimensions to obtain $(n^d \cdot \# Q_n)$ subcubes with sidelengths $1/n$. For each cube in Q'_{n+1} we then decide independently with probability p to keep the cube. We set Q_{n+1} to be the set of 'surviving' cubes.

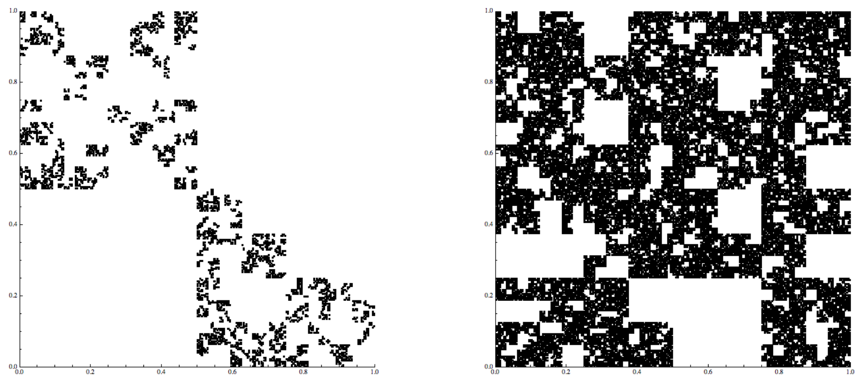

Figure 1: Mandelbrot percolation for $p = 0.7$ and $p = 0.9$ ($n = 2$, $d = 2$).

First, we compute the probability generating function that determines how many cubes we keep in the first division step. It is of binomial distribution, but we could equally let X be the random variable such that $\mathbb{P}\{X = 1\} = p$ and $\mathbb{P}\{X = 0\} = 1 - p$. Its probability generating function is $f(X, s) = 1 - p + p\,s$. Thus the probability generating function of Y, the number of subcubes we keep at every division step is

$$f(Y, s) = f\Big(\sum_{i=1}^{n^d} X_i, s\Big) = \mathbb{E}(s^{\sum_{i=1}^{n^d} X_i}) = \mathbb{E}(s^X)^{n^d} = f(X, s)^{n^d} = (1 + p(s-1))^{n^d}.$$

Since the subdivision process is easily seen to be a Galton–Watson process with offspring distribution given by $\theta_i = \mathbb{P}\{Y = i\}$, the probability of extinction is the least non-negative q that satisfies $(1+p(q-1))^{n^d} = q$. We can further determine the threshold p_0 such that for $0 < p \leq p_0$ we have almost sure extinction and for $p_0 < p \leq 1$ there is positive probability that the Mandelbrot percolation is non-empty. Using Theorem 1.7 yet again the threshold is for p_0 satisfying $m = f^{(1)}(Y, 1) = 1$. Differentiating, with respect to s,

$$f^{(1)}(Y, s) = n^d p(1 - p + ps)^{n^d - 1}$$

and so p_0 satisfies $1 = n^d p_0 (1 - p_0 + p_0\,1)^{n^d-1} = n^d p_0$. That is, $p_0 = 1/n^d$. Further, our computations have shown that $m = pn^d$ and we will come back to this number when analysing the box-counting dimension of Mandelbrot percolation.

2 Martingales and Convergence Theorems

In this section we define the discrete-time *martingale* and state two convergence results. Assume we are given a stochastic process, *i.e.* a sequence of random variables, $(X_i)_i$. Informally, a *martingale* is a process where, given information about all previous outcomes, the expectation of the next outcome is the current value.

For martingales we will also need the notion of *conditional expectation*.

Definition 2.1 (Conditional expectation)**.** *Let X be a random variable and $\mathscr{A}' \subseteq \mathscr{A}$ be a σ-algebra. The* conditional expectation of X given $\mathscr{A}'$, *denoted by $\mathbb{E}(X|\mathscr{A}')$, is any $\mathscr{A}'$-measurable function $\Omega \to \mathbb{R}$ satisfying*

$$\int_{A'} \mathbb{E}(X|\mathscr{A}')d\,\mathbb{P} = \int_{A'} X d\,\mathbb{P}$$

for all $A' \in \mathscr{A}'$.

The conditional expectation can be interpreted as the expectation with the 'knowledge' of events in the σ-algebra $\mathscr{A}'$. The 'finer' the σ-algebra, the better our prediction of the outcome. As an example, if $\mathscr{A}' = \{\varnothing, \Omega\}$ we have no knowledge and the conditional expectation is a constant function $\mathbb{E}(X|\mathscr{A}') = \mathbb{E}(X)$. However, if $\mathscr{A}' = \mathscr{A}$, we have 'total knowledge' and $\mathbb{E}(X|\mathscr{A}') = X$. Given a random variable, we define $\langle X \rangle$, the σ-algebra generated by X, to be the smallest σ-algebra such that X is Borel measurable.

It is important to note that $\mathbb{P}$ is a measure on the event space Ω. Given a fixed r.v. we might instead want to talk about the *distribution* of measurements. The *distribution* $\mathbb{P}_X$ is simply the image of the measure $\mathbb{P}$ under X, *i.e.* $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$ for all $A \subset \mathbb{R}$, which is itself a measure.

Definition 2.2. *A discrete stochastic process $(M_i)_i$ is called a* (discrete-time) martingale *if all M_i are integrable and*

$$\mathbb{E}(M_{i+1} \mid \langle M_1, M_2, \ldots, M_i \rangle) = M_i \quad \text{for all } i.$$

Equivalently,

$$\mathbb{E}(M_{i+1} - M_i \mid \langle M_1, M_2, \ldots, M_i \rangle) = 0 \quad \text{for all } i.$$

Similarly, a stochastic process is called a submartingale *if it satisfies*

$$\mathbb{E}(M_{i+1} \mid \langle M_1, M_2, \ldots, M_i \rangle) \geq M_i \quad \text{for all } i,$$

and a supermartingale *if*

$$\mathbb{E}(M_{i+1} \mid \langle M_1, M_2, \ldots, M_i \rangle) \leq M_i \quad \text{for all } i.$$

We note that every martingale is also a supermartingale and submartingale.

2.1 Doob's Martingale convergence theorems

The first convergence result we mention is due to Doob about pointwise convergence and a proof can be found in [Wil91].

Theorem 2.3. *Let (M_i) be a supermartingale in $\mathcal{L}^1$, that is $\sup_i \mathbb{E}(|M_i|) < \infty$. Then the pointwise limit $M = \lim_{i\to\infty} M_i$ exists almost surely and $\mathbb{E}(M) < \infty$.*

Corollary 2.4 (Non-negative Martingale Convergence Theorem[2])**.** *Let (M_i) be a non-negative supermartingale. Then the pointwise limit $M = \lim_{i\to\infty} M_i$ exists almost surely.*

[2] The name of this theorem is also the only theorem known to the author for which the number of words in English necessary to state the theorem adequately is no greater than the number of words in its title. *The non-negative martingale convergence theorem: non-negative martingales converge.*

Proof. Since $\mathbb{E}(|M_i|) = \mathbb{E}(M_i) \leq \mathbb{E}(M_0)$, the process is in $\mathcal{L}^1$ and we can apply the pointwise Martingale convergence theorem. $\square$

Example 2.5. *Let S_0 be my accumulated wealth and assume that it is positive and finite. Assume further, that I have quit my job and will not have any income anymore from any source. Also, assume that I have no expenditures and do not consume food of any kind. I turn to the only thing left: gambling. Note that, even though we do not know what game I am playing, we can say that the games are at best fair games and my wealth after the k-th game only depends on the wealth I have had before. In other words, $\mathbb{E}(S_k \mid \langle S_{k-1} \rangle) \leq S_{k-1}$. Because of my bad credit history, I am not eligible for any credit. Thus, assume that $S_k \geq 0$ for all k,* i.e. *I stop playing if I run out of money. My wealth therefore is a non-negative martingale and we can apply Theorem 2.3 and conclude that my wealth converges almost surely, independent of any gambling strategy, almost surely I will stop playing and end up with a fixed amount of wealth. Knowing my habits, $S_k = 0$, for some big enough k.*

We note that these two results do not imply convergence in $\mathcal{L}^1$, *i.e.* $\mathbb{E}(|M_i - M|)$ may not tend to 0 as $i \to \infty$. We will shortly see an illuminating example of this but if we want this convergence, or uniform convergence, we need stronger assumptions and state another result by Doob, a proof of which can also be found in [Wil91].

Theorem 2.6. *Let (M_i) be a supermartingale such that $\sup_i \mathbb{E}|M_i|^p < \infty$ for some $p > 1$. Then there exists a random variable $M \in \mathcal{L}^p(\Omega, \mathbb{R})$ such that*

$$M_i \to_{a.s.} X \qquad \text{and} \qquad \int_\Omega |M_i - M|^p d\,\mathbb{P} \to 0.$$

In particular, $\mathbb{E}(|M_i|^p) \to \mathbb{E}(|M|^p)$.

2.2 Application to Galton–Watson processes

The first of the two convergence theorems is useful if you only require almost sure convergence and do not need to know anything about the limit. As an example, we look at a process that we can obtain from the Galton–Watson process, called the *normalised Galton–Watson process.*

Definition 2.7. *Let $\vec{\theta}$ be a non-trivial offspring distribution with associated Galton–Watson process (Z_k) with finite mean, $m = \mathbb{E}(Z_0)$. The* normalised Galton–Watson process *is*

$$W_k = Z_k / m^k.$$

We immediately see that $\mathbb{E}(W_k) = 1$ for all $k > 0$.

Lemma 2.8. *The stochastic process (W_i) is a non-negative martingale and thus converges almost surely to some random variable W with finite expectation.*

Proof. Since $Z_n \geq 0$ and by non-triviality the mean m is positive, we also have $W_i \geq 0$ for all i. It remains to check that (W_k) is a martingale.

$$\begin{aligned}
\mathbb{E}(W_k \mid \langle W_1, \dots, W_{k-1} \rangle) &= \mathbb{E}(Z_k/m^k \mid \langle Z_1, \dots, Z_{k-1} \rangle) \\
&= \mathbb{E}(Z_k/m^k \mid \langle Z_{k-1} \rangle) \\
&= \mathbb{E}\left(1/m^k \sum_{i=1}^{Z_{k-1}} X_k^i \;\middle|\; \langle Z_{k-1} \rangle\right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left(1/m^k Z_{k-1} \;\middle|\; \langle Z_{k-1}\rangle\right) \mathbb{E}(X_k^i) \\
&= (Z_{k-1}/m^{k-1})\, \mathbb{E}(Z_0)/m = W_{k-1},
\end{aligned}
$$

as required. The convergence follows from Corollary 2.4. □

This convergence allows us to conclude that almost every Galton Watson process satisfies $C_\omega^{-1} m^k \leq Z_k \leq C_\omega m^k$ for some random C_ω and big enough k. However, $C_\omega = 0$ almost surely, is not excluded and would give a fairly trivial result. We note further, that such a Galton–Watson process may not even converge in $\mathcal{L}^1$. Let $\vec{\theta}$ be such that its mean satisfies $m = 1$. We have already learnt that $Z_n = 0$ for big enough n is an almost sure event. Its normalised Galton–Watson process (W_n) then coincides with (Z_n) and both have unit expectation for all n. But almost sure extinction means that $\mathbb{E}(W) = 0$ and so

$$
\mathbb{E}(|W_n - W|) = \mathbb{E}(|Z_n - Z|) = \mathbb{E}(Z_n) = 1 \not\to 0.
$$

When it comes to $\mathcal{L}^p$ convergence, the easiest case to check is usually $p = 2$, since

$$
\mathbb{E}(M_n^2) = \mathbb{E}(M_0^2) + \sum_{i=1}^{n} \mathbb{E}((M_i - M_{i-1})^2).
$$

However, for Galton–Watson processes there is an even more convenient way to determine the second moment. Let $\vec{\theta}$ be any offspring distribution, we observe that

$$
f^{(2)}(X, s) = \sum_{i=2}^{\infty} \theta_i (i)(i-1) s^{i-2} = \sum_{i=0}^{\infty} \theta_i (i)(i-1) s^{i-2} = \sum_{i=0}^{\infty} \theta_i\, i^2\, s^{i-2} - f^{(1)}(X, s).
$$

This holds in general and we find

$$
\begin{aligned}
\mathbb{E}((Z_k - m^k)^2) &= \sum_{i=0}^{\infty} \mathbb{P}\{Z_k = i\}(i - m^k)^2 \\
&= \sum_{i=0}^{\infty} \mathbb{P}\{Z_k = i\}\, i^2 + \sum_{i=0}^{\infty} \mathbb{P}\{Z_k = i\}(m^k)^2 - 2 \sum_{i=0}^{\infty} \mathbb{P}\{Z_k = i\}\, i\, m^k \\
&= \sum_{i=0}^{\infty} \mathbb{P}\{Z_k = i\}\, i^2 + (m^k)^2 - 2m^k f(Z_k, 1) \\
&= \sum_{i=0}^{\infty} \mathbb{P}\{Z_k = i\}\, i^2 - m^{2k} \\
&= f_k^{(2)}(X, 1) + f_k^{(1)}(X, 1) - m^{2k} \\
&= f_{k-1}^{(2)}(X, f(X, 1))(f^{(1)}(X, 1))^2 + f_{k-1}^{(1)}(X, f(X, 1)) f^{(2)}(X, 1) + m^k - m^{2k} \\
&= f_{k-1}^{(2)}(X, 1) m^2 + f_{k-1}^{(1)}(X, 1) f^{(2)}(X, 1) + m^k - m^{2k} \\
&= f_{k-1}^{(2)}(X, 1) m^2 + m^{k-1} f^{(2)}(X, 1) + m^k - m^{2k} \\
&= (f_{k-2}^{(2)}(X, 1) m^2 + m^{k-2} f^{(2)}(X, 1)) m^2 + m^{k-1} f^{(2)}(X, 1) + m^k - m^{2k} \\
&= f_{k-2}^{(2)}(X, 1) m^4 + f^{(2)}(X, 1)(m^{k-1} + m^k) + m^k - m^{2k} \\
&= f_{k-3}^{(2)}(X, 1) m^6 + f^{(2)}(X, 1)(m^{k-1} + m^k + m^{k+1}) + m^k - m^{2k} \\
&= f^{(2)}(X, 1)(m^{k-1} + m^k + m^{k+1} + \cdots + m^{2k-2}) + m^k - m^{2k}
\end{aligned}
$$

Now $\operatorname{Var}(X) = f^{(2)}(X,1) + m - m^2$ and so

$$\operatorname{Var}(Z_k) = \mathbb{E}((Z_k - m^k)^2) = \operatorname{Var}(X) m^{k-1}(1 + m + \cdots + m^{k-1}) = \operatorname{Var}(X) m^{k-1} \frac{m^k - 1}{m - 1}.$$

So,

$$\operatorname{Var}(W_k) = \left(\frac{1}{m^k}\right)^2 \operatorname{Var}(Z_k) = \operatorname{Var}(X) \frac{m^k - 1}{m^k\,(m^2 - m)} = \operatorname{Var}(X) \frac{1 - m^{-k}}{m(m-1)},$$

which is bounded above for $m > 1$. This means that we can apply the second martingale convergence theorem and there exists $W \in \mathcal{L}^2$ such that $W_k \to W$ almost surely. Moreover, since W_k is non-negative, $1 = \mathbb{E}(W_k) \to \mathbb{E}(W)$ and thus there exists positive probability that $W_k \asymp m^k$. Further, since $\mathbb{P}\{W = \infty\} = 0$, we can claim that W_k converges to some positive real number almost surely, conditioned on non-extinction.

Box-counting dimension of Mandelbrot percolation This can now be applied to calculate the *box-counting dimension* of Mandelbrot percolation.

Definition 2.9. *Let $(\mathbf{M}, d)$ be a totally bounded metric space. Denote by $N_\delta(\mathbf{M})$ the least number of sets of diameter $\delta > 0$ needed to cover $\mathbf{M}$. The box-counting dimension of $\mathbf{M}$ is defined as*

$$\dim_B \mathbf{M} = \lim_{\delta \to 0} \frac{\log N_\delta(\mathbf{M})}{-\log \delta}.$$

If the limit does not exist we talk about upper and lower box-counting dimension, taking the upper and lower limit, respectively.

Assume that $p > n^d$, so that there is a positive probability that the limit set is non-empty. One can easily see that the least number of sets of diameter $\delta = n^{-k}$ is comparable to the number of subcubes at stage k. Since those are Galton–Watson processes with bounded variance, the above result can be applied to give $\#Q_k \asymp m^k$. Thus

$$\frac{\log N_{n^{-k}}(M(\omega))}{-\log n^{-k}} \leq \frac{\log C_\omega m^k}{k \log n} = \frac{\log C_\omega}{k \log n} + \frac{\log(p\, n^d)}{\log n}$$

for some C_ω that is almost surely positive and finite, conditioned on non-extinction. A similar lower bound holds and upon taking limits we conclude that the almost sure box-counting dimension of Mandelbrot percolation is $\log(n^d p)/\log n$. The keen observer might note that we only looked at $\delta = n^{-k}$, however, this does not produce any issues as the number of subcubes surviving is off by at most a fixed multiple to the least number of covering sets. We leave details to the reader.

2.2.1 Random Cascade Measure

The same process can be applied to labelled trees to generate a random measure on $[0,1]^d$ called the *random cascade measure*. Let X be a non-negative random variable with mean $\mathbb{E}(X) = n^{-d}$ and consider a process similar to the Mandelbrot percolation. At each step we subdivide a subcube into n^d subcubes but instead of deleting cubes we associate a realisation of X_C with each subcube C. Let $C \subset Q_k$ be a level k cube with address $C_1 C_2 \ldots C_k$. We define its l-level mass by

$$m_l(C) = X_{C_1} X_{C_2} \ldots X_{C_k} \sum_{\substack{D \in Q_{k+l} \\ D_i = C_i \\ 1 \leq i \leq k}} D_{k+1} D_{k+2} \ldots D_{k+l}$$

Now $m_l(C)$ is a non-negative martingale and the pointwise convergence theorem tells us that m_l converges pointwise and it is easily verifiable that the limit m is indeed a measure on the unit cube. This random measure is the *random cascade measure*.

Let $d = 1$ and $n = 2$. That is, the subcubes are the dyadic intervals. Let X be a random variable with mean $1/2$ and $\mathbb{P}\{X = 0\} = 0$. Consider the associated random cascade measure and assume $\mathbb{E}(X \log X) < \infty$. It can be shown that the associated random cascade measure μ of the unit interval is almost surely positive. Without loss of generality we assume $\mu([0,1]) = 1$ and set $\phi(x) = \mu([0,x))$ to be the cumulative probability distribution function and note that it is strictly increasing (almost surely).

Now let F be any subset of the unit line, a question that is currently of active interest is what happens to the dimensions of F under the mapping ϕ, *i.e.* what is the relationship between $\dim F$ and $\dim \phi(F)$.

3 Finer Information on the number of descendants

So far, we have learnt about the mean of the Galton–Watson process at step n, and achieved finer almost sure results using martingales. In this section we will go back to the probability generating function to prove a Lemma due to Athreya [Ath94]. This Lemma will allow us to estimate the number of descendants at a given node and we will apply this to establish a similar result as in [Tro19] but for Mandelbrot percolation. process.

We first introduce additional notation. Notice that f, the generating function, was convex and strictly increasing. Hence it is invertible with strictly increasing but concave derivative. We denote the inverse map of f by g and write g_k for the k-th iterate of the inverse.

From now on we make an assumption that is somewhat stronger than the mean existing, namely that the generating function is defined on an interval greater than $[0,1]$. We generally make the assumption that there exists $s_0 > 1$ such that $f(s_0) < \infty$. This implies that f is defined on $[0, s_0]$ and g on $[f(0), f(s_0)] = [\theta_0, f(s_0)]$. As a side-note, this further means that the mean is defined since $s_0 > 1$ forces $f^{(1)}(1) < \infty$.

Lemma 3.1. *Let $f(s_0) < \infty$ for some $s_0 > 1$. Then, for $1 \le s \le f(s_0)$ and $g_k(s_0) \searrow 1$ from above,*

$$m^k(g_k(s) - 1) \searrow G(s),$$

where $G(s)$ is the unique solution to $G(f(s)) = mG(s)$ for $1 \le s \le f(s_0)$. Further $0 < G(s) < \infty$ for all $1 < s \le f(s_0)$ and $G(1) = 0$ and $G'(1) = 1$.

The proof is similar to the ones we have seen for the generating function and we omit details.

Lemma 3.2 ([Ath94, Theorem 4])**.** *Let Z_k be a Galton–Watson process with mean $m = \mathbb{E}(X^\varepsilon) < \infty$. Suppose that there exists $t_0 > 0$ such that $\mathbb{E}(\exp(t_0 Z_1) \mid Z_0 = 1) < \infty$. Then there exists $t_1 > 0$ such that*

$$\sup_{k \in \mathbb{N}} \mathbb{E}\left(e^{t_1 W_k} \,\middle|\, Z_0 = 1\right) < \infty \tag{3.1}$$

Proof. By assumption, there exists some $s_0 = e^{t_0} > 1$ such that $f(s_0) < \infty$. Let $K = f(s_0) < \infty$, we must have $f_2(s) \le K$ if $0 \le f(s) \le s_0$. Equivalently, $0 \le s \le g(s_0)$ and in general we get

$$f_n(s) \le K \quad \text{if} \quad 0 \le s \le g_{n-1}(s_0).$$

Recall that W_k was the normalised Galton–Watson process and so

$$\mathbb{E}\left(e^{tW_k} \;\middle|\; Z_0 = 1\right) = f_n\left(e^{tm^{-k}}\right).$$

Thus,

$$\mathbb{E}\left(e^{tW_k} \;\middle|\; Z_0 = 1\right) \le K \quad \text{if} \quad t \le m^k \log g_{n-1}(s_0).$$

It is a simple argument, similar to Theorem 1.7, to establish that $g_k(t) \to 1$ for all $t > 1$ where g is defined. We can therefore conclude that $\log g_k(s_0) \sim g_k(s_0) - 1$ and make use of Lemma 3.1, and have

$$m^k \log g_{k-1}(s_0) \to mG(s_0),$$

which is positive and finite. Since, further $g_k(s_0) > 1$ for all $k \ge 1$ we can choose

$$t_1 = \inf_k m^k \log g_{k-1}(s_0)$$

and the left-hand side of (3.1) is bounded by K. □

With this technical lemma out of the way we can now focus our attention on an interesting consequence with regards to the deviation from the mean number of children.

Theorem 3.3. *Let Z_k be a Galton–Watson process with non-trivial offspring distribution. Assume that there exists t_0 such that $\mathbb{E}(\exp(t_0 Z_1) \mid Z_0 = 1) < \infty$. Let $C > 0$ and $\varepsilon > 0$ be given. Then there exist $t_2 > 0$ and $D > 0$ such that*

$$\mathbb{P}\left\{Z_k \ge Cm^{(1+\varepsilon)k}\right\} \le De^{-t_2 m^{\varepsilon k}}.$$

That is, the probability that Z_k exceeds $Cm^{(1+\varepsilon)k}$ decreases superexponentially.

Proof. Let t_1 and K be given by Lemma 3.2. We use a standard Chernoff bound to obtain,

$$\begin{aligned}
\mathbb{P}\left\{Z_k \ge Cm^{(1+\varepsilon)k}\right\} &= \mathbb{P}\left\{W_k \ge Cm^{\varepsilon k}\right\} \\
&= \mathbb{P}\left\{\exp(t_1 W_k) \ge \exp\left(Ct_1 m^{\varepsilon k}\right)\right\} \\
&\le \frac{\mathbb{E}(\exp(t_1 W_k))}{\exp(Ct_1 m^{\varepsilon k})} \\
&= Ke^{t_2 m^{\varepsilon k}},
\end{aligned}$$

for $t_2 = Ct_1$, as required. □

We can state an immediate corollary.

Corollary 3.4. *Let Z_k be a Galton–Watson process satisfying the conditions of Theorem 3.3. Let $C > 0$ and $\varepsilon > 0$ be given. Then,*

$$\mathbb{P}\left\{Z_k \ge Cm^{(1+\varepsilon)k} \text{ for some } k \ge l\right\} \le \sum_{k=l}^{\infty} De^{-t_2 m^{\varepsilon k}} \le D' e^{-t_2 m^{\varepsilon l}},$$

for some $D' > 0$ independent of l.

3.1 Application: Mandelbrot percolation (yet again)

In fractal geometry there are several notions of dimension that have served, and will serve, as the main field we apply our results to. Apart from the box-counting dimension introduced in Section 2.2, we now consider the Assouad and quasi-Assouad dimension. Write $N_{r,R}(X) = \max_{x\in X} N_r\left(B\left(x,R\right)\cap F\right)$ for the minimal number of centred open r balls needed to cover any open ball of F of diameter less than R. Let

$$h(\delta,F) = \inf\left\{\alpha \geq 0 \mid \exists C>0, \forall 0<r\leq R^{1+\delta}\leq \operatorname{diam}(F) \text{ we have } N_{r,R}(F)\leq C(R/r)^{\alpha}\right\}$$

The Assouad dimension is given by $\dim_A(F) = h(0,F)$; it is the minimal exponent such that all open balls of F can be covered by a certain number of r balls relative to the size of the ball of F. The Assouad dimension is an important tool in the study of embeddings by bi-Lipschitz maps. We refer to Fraser [Fra14] for more information on the Assouad dimension.

We note that $\delta = 0$ gives no restriction on the ratio R/r other than that it is greater than one. For positive $\delta > 0$ this means, however, that there must be a gap between r and R that grows as R decreases. Further, $h(\delta,F)$ may not be continuous in δ at zero, as was shown by García and Hare [GH17], and we call this limit the quasi-Assouad dimension $\dim_{qA}(F) = \lim_{\delta\to 0} h(\delta,F)$.

We can easily determine the almost sure Assouad dimension of Mandelbrot percolation.

Theorem 3.5. *Let M be the limit set of Mandelbrot percolation with parameter $p > n^{-d}$. Conditioned on non-extinction, almost surely,*

$$\dim_A(M) = d.$$

We note that it is maximal and independent of p and provide a sketch of the proof below.

Proof. Let $N = n^d$. As we have learnt from Section 2.2, there exists probability $0 < q_0 \leq 1$ such that any subcube has at least one surviving descendent. Therefore the probability that there exists a full subtree for k levels such that there exists at least one descendent in every last subcube is $p^N p^{N^2} p^{N^3} \dots p^{N^k} q_0^{N^{k+1}} = p^{l(k)} q_0^{N^{k+1}}$, for an appropriate $l(k)$. Let $L(k)$ be the least integer such that

$$\left(1 - p^{l(k)} q_0^{N^{k+1}}\right)^{L(k)} < 1/2.$$

This means that the probability such that there exists at least one k block in $L(k)\,k$ levels is at least $1/2$. We can thus partition the infinite levels in chunks: the first from level 1 to $L(1)$, then from $L(1)+1$ to $2L(2)$, et cætera. Each of these blocks are independent and have probability $1/2$ of containing a full tree of length k. Applying Borel-Cantelli, this must happen infinitely often with full probability and hence, almost surely, there are infinitely many full subtrees of arbitrary length. Letting $B(x,R)$ be comparable to the start of such blocks, we conclude that we require roughly n^{kd} many $r = n^{-k}R$ balls to cover $B(x,R)$. This gives the desired conclusion. □

Surprisingly, the quasi-Assouad dimension is as minimal as it can be: It coincides with the almost sure box-counting dimension.

Theorem 3.6. *Let M be the limit set of Mandelbrot percolation with parameter $p > n^{-d}$. Conditioned on non-extinction, almost surely,*

$$\dim_{qA}(M) = \dim_B(M) = \frac{\log pn^d}{\log n}.$$

Again, we only sketch the proof, a full version of which can be found in [Tro19].

Proof. Let $\delta > 0$ and assume for a contradiction that there exists $\varepsilon > 0$ such that $\dim_{qA}(M) \geq \dim_B(M) + \varepsilon$ with positive probability. Thus we should be able to find a sequence of balls $B(x, R)$ such that the minimal number of covering sets exceeds $(pn^d)^{(1+\varepsilon)k}$, where $k \geq \log R / \log r$. This is equivalent to finding infinitely many subcubes in the Mandelbrot percolation at levels l_k with more than $(pn^d)^{(1+\varepsilon)}$ subcubes at levels exceeding $l_k^{1+\delta'}$ for some $\delta' > 0$ only dependent on δ. We know from Corollary 3.4 that the probability of this occurring decreases superexponentially. However, the number of descendants grows at most exponentially since they are bounded by n^d children at every step and we get

$$\mathbb{P}\left\{\exists k\text{-level subcube with "too many" descendants}\right\} \leq \left(n^d\right)^k D' e^{-t_2 m^{\varepsilon(1+\delta')k}}$$

and we note that summing the right-hand term over k is finite. But then the probability that there are infinitely many levels that contain such a full enough subtree is zero by the Borel–Cantelli Lemmas, contradicting the positive measure assumption. □

4 Kingman's Subadditive Ergodic Theorem

Let (Ω, μ) be a probability space. Consider the surjective map $T : \Omega \to \Omega$. We assume that T is invariant with respect to μ, *i.e.* $\mu(A) = \mu(T^{-1}A)$ for all measurable $A \subseteq \Omega$. Let $F : \Omega \to \mathbb{R}$ be a real-valued measurable function. We say that T is *ergodic* with respect to μ if $A = T^{-1}A$ implies $\mu(A) = 0$ or $\mu(\Omega \setminus A) = 0$ for all measurable $A \subseteq \Omega$.

The famous *Birkhoff's Ergodic Theorem* then states that for μ-almost every $\omega \in \Omega$,

$$\frac{1}{n}\sum_{j=1}^{n} F(T^j(\omega)) \to \int f \, d\mu,$$

if T is ergodic. Informally, we say that the *time average* converges to the *space average* for *observable* F.

4.1 Fekete's Lemma

To motivate Kingman's subadditive theorem, consider the following well known result on sub-additive sequences. Let $(a_i)_{i=1}^{\infty}$ be a real-valued sequence such that $a_{n+m} \leq a_n + a_m$. We call any such sequence *subadditive.* Fekete's Lemma states that a_n/n converges if $(a_i)_{i=1}^{\infty}$ is subadditive.

Lemma 4.1 (Fekete's Lemma)**.** *Let $(a_i)_{i=1}^{\infty}$ be a subadditive sequence. Then,*

$$\frac{a_n}{n} \to \liminf_{i\to\infty} \frac{a_i}{i} \in [-\infty, \infty).$$

While the proof does not help us for Kingman's subadditive ergodic theorem, it is straightforward and we include it for completeness.

Proof. Let $q \in \mathbb{N}$. Then, for every n, there exists a unique $p_n \in \mathbb{N}$ and $r_n \in \{0, 1, \ldots, p-1\}$ such that $n = pq + r$. Therefore, using subadditivity,

$$\frac{a_n}{n} = \frac{a_{p_n q + r_n}}{p_n q + r_n}$$

$$
\begin{aligned}
&\leq \frac{a_{p_n q} + a_{r_n}}{p_n q} \\
&\leq \frac{p_n a_q + a_{r_n}}{p_n q}
\end{aligned}
$$

and therefore, taking the upper limit,

$$
\limsup_{n\to\infty} \frac{a_n}{n} \leq \limsup_{n\to\infty} \frac{p_n a_q + a_{r_n}}{p_n q} = \frac{a_q}{q}.
$$

But this holds for any $q \in \mathbb{N}$ and so, in particular, $\liminf_{i\to\infty} a_i/i$ is a (finite) upper bound. □

4.2 Kingman's Subadditive Ergodic Theorem (at last)

In this section we state and prove Kingman's subadditive theorem. The original theorem is due to Kingsman [Kin73] and the short proof we present is due to Steele [Ste89].

Theorem 4.2. *Let T be a μ-invariant surjective transformation on Ω and let g_n be a sequence of integrable sub-additive functions,*

$$
g_{n+m}(\omega) \leq g_n(\omega) + g_m(T^n\omega). \tag{4.1}
$$

Then, for μ-almost every $\omega \in \Omega$,

$$
\lim_{n\to\infty} \frac{g_n(\omega)}{n} = g(\omega) \geq -\infty,
$$

where g is an invariant function. Further, if T is also ergodic, then g is constant almost surely.

Proof. Instead of g_n we could consider the function

$$
\hat{g}_n(\omega) = g_n(\omega) - \sum_{i=1}^{n-1} g_1(T^i\omega). \tag{4.2}
$$

Using the subadditivity this implies $\hat{g}_n \leq 0$ and further $\hat{g}_n$ satisfies the same subadditivity condition of equation 4.1. As we are interested in g_n/n, we see that the second term in (4.2), when normalised by n, converges for almost every ω due to Birkhoff's Ergodic Theorem. Therefore, g_n/n converges if and only if $\hat{g}_n/n$ converges. We can therefore, without loss of generality, assume that g_n is non-positive.

For definitiveness, we define $g(\omega) = \liminf_{n\to\infty} g_n(\omega)/n$. We observe that g is an invariant function with respect to T. Since $g_{n+1}(\omega)/n \leq g_1(\omega)/n + g_n(T(\omega))/n$ we take the lower limit and obtain $g(\omega) \leq g(T(\omega))$. But then

$$
\{\omega \in \Omega \mid g(\omega) \geq \alpha\} \subseteq \{\omega \in \Omega \mid g(T(\omega)) \geq \alpha\} = T^{-1}\{\omega \in \Omega \mid g(\omega) \geq \alpha\}
$$

and so the sets on the left-hand side and right-hand differ by at most a set of zero measure, since T is measure preserving. This means that $g(\omega) = g(T(\omega))$ almost surely. Further, if T is ergodic with respect to μ, the intersection of the two sets must have measure zero or one. But then there exists some "jump value" $\alpha_0 \in [-\infty, \infty)$ for which the measure of the set above jumps from 1 to 0. Therefore $g(\omega) = \alpha_0$ almost surely. By taking countable intersections of full measure subsets of Ω we can thus, without loss of generality, assume that $g(T^k(\omega)) = g(\omega)$ for all k.

To get the required convergence results we will employ a few tricks. The first one concerns the lower bound. Nothing is preventing $g = -\infty$ for some, or all $\omega \in \Omega$ (nor did we claim so), and we will circumvent the slight issue this is presenting by defining $G_b(\omega) = \min\{-b, g(\omega)\}$ for $b \geq 0$. Now, let $\varepsilon > 0$ and $b \geq 0$ be given. For every $N \in \mathbb{N}$ we define the *abysmal* words

$$A_{b,\varepsilon}(N) = \{\omega \in \Omega \mid g_l(\omega)/l > G_b(\omega) + \varepsilon \text{ for all } 1 \leq l \leq N\}.$$

Correspondingly we let $S_{b,\varepsilon}(N) = \Omega \setminus A_{b,\varepsilon}(N)$ be the set of *splendid* words.

Consider an arbitrary $\omega \in \Omega$ and let $n > 1$ be given. We consider the integer set $\{1, 2, 3, \ldots, n-1\}$ and will decompose it into several disjoint sets of three different types: τ_1, τ_2, and τ_3. This follows the following algorithm.

1. Set $k = 1$.
2. Consider $\omega' = T^k(\omega)$:
   (a) If $\omega' \in A_{b,\varepsilon}(N)$ add the singleton set $\{k\}$ to τ_1. Go to Step 3.
   (b) If $\omega' \in S_{b,\varepsilon}(N)$, there exists l such that
   $$g_l(T^k(\omega)) \leq l(G_b(T^k(\omega)) + \varepsilon) = l(G_b(\omega) + \varepsilon).$$
   If there are multiple, consider the least l and evaluate $l + k$:
      i. If $l + k < n$, add the set $\{k, k+1, \ldots, k+l-1\}$ to τ_2. Go to Step 3.
      ii. If $l + k \geq n$, add the singleton set $\{k\}$ to τ_3. Go to Step 3.
3. Check whether there are any integers in $\{1, \ldots, n-1\}$ not contained in any of τ_1, τ_2, or τ_3. If there is, let k be the least such integer and go to Step 2. Otherwise, terminate the algorithm.

We write $\#\tau_j$ for the number of sets in class τ_j and write l_i for the number of elements in each of the sets. Further, we denote by k_i the least element in each of these sets. Note that we can use the subadditivity to decompose $g_n(\omega)$ as the sum of values at each of the pairs determined above, that is

$$g_n(\omega) \leq \sum_{i=1}^{\#\tau_1} g_1(T^{k_i}(\omega)) + \sum_{i=1}^{\#\tau_2} g_{l_i}(T^{k_i}(\omega)) + \sum_{i=1}^{\#\tau_3} g_1(T^{k_i}(\omega)) \leq \sum_{i=1}^{\#\tau_2} g_{l_i}(T^{k_i}(\omega)),$$

where the last inequality holds as $g_n(\omega) \leq 0$. But since we know that $T^{k_i}(\omega)$ is splendid we can further write

$$g_n(\omega) \leq \sum_{i=1}^{\#\tau_2} l_i(G_b(\omega) + \varepsilon) \leq G_b(\omega) \sum_{i=1}^{n} l_i + n\varepsilon, \tag{4.3}$$

where we take $l_i = 1$ for all $i > \#\tau_2$. Note further that by virtue of construction

$$n \leq \sum_{i=1}^{\#\tau_2} l_i + \sum_{i=1}^{n} \chi_{A_{b,\varepsilon}(N)}(T^i(\omega)) + N$$

and so,

$$\frac{1}{n}\sum_{i=1}^{\#\tau_2} l_i \geq 1 - \frac{N}{n} - \frac{1}{n}\sum_{i=1}^{n} \chi_{A_{b,\varepsilon}(N)}(T^i(\omega))$$

Taking the lower limit and applying Birkhoff's Ergodic Theorem we obtain, for almost every $\omega \in \Omega$,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} l_i \geq 1 - \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\chi_{A_{b,\varepsilon}(N)}(T^i(\omega)) = 1 - \mathbb{E}\left(\chi_{A_{b,\varepsilon}(N)}\right) = 1 - \mu\left(A_{b,\varepsilon}(N)\right)$$

Combining this with the upper bound in (4.3) we get, almost surely,

$$\limsup_{n\to\infty}\frac{g_n(\omega)}{n} \quad \leq \quad G_b(\omega)(1-\mu(A_{b,\varepsilon}(N))+\varepsilon \quad \leq \quad G_a(\omega)(1-\mu(A_{b,\varepsilon}(N))+\varepsilon$$

for all $a \leq b$.

We are now almost done. First letting $N \to \infty$ gives

$$\limsup_{n\to\infty} g_n(\omega)/n \leq G_a(\omega)+\varepsilon,$$

which holds for all $a \geq 0$ and $\varepsilon > 0$. Thus,

$$\limsup_{n\to\infty} g_n(\omega)/n \leq \liminf_{n\to\infty} g_n(\omega)/n = g(\omega),$$

almost surely, as required. □

4.3 An Application

Let $\mathbf{M}_\lambda$ be $n \times n$ matrices, indexed by some $\lambda \in \Lambda$. Let μ be a probability measure supported on Λ. Let $\Omega = \Lambda^{\mathbb{N}}$ be the set of infinite sequences over Λ and set $\mathbb{P} = \mu^{\mathbb{N}}$. It is easy to check that $\mathbb{P}$ is invariant with respect to $T(\omega) = T(\omega_1\omega_2\dots) = \omega_2\omega_3\dots$. Slightly less trivial, the measure $\mathbb{P}$ is also ergodic with respect to T. Let $\|.\|$ be any submultiplicative matrix norm. Then,

$$\|\mathbf{M}_{\omega_1}\mathbf{M}_{\omega_2}\dots\mathbf{M}_{\omega_k}\|^{1/k} \to c \in [0,\infty) \quad \text{as} \quad k\to\infty \quad \text{(a.s.)}. \tag{4.4}$$

This can quickly be proved by letting $g_n(\omega) = \log\|\mathbf{M}_{\omega_1}\mathbf{M}_{\omega_2}\dots\mathbf{M}_{\omega_n}\|$ which is subadditive by the submultiplicativity of the norm.

$$\begin{aligned} g_{n+m}(\omega) &= \log\|\mathbf{M}_{\omega_1}\mathbf{M}_{\omega_2}\dots\mathbf{M}_{\omega_{n+m}}\| \\ &\leq \log\|\mathbf{M}_{\omega_1}\dots\mathbf{M}_{\omega_n}\| + \log\|\mathbf{M}_{\omega_{n+1}}\dots\mathbf{M}_{\omega_{n+m}}\| \\ &= g_n(\omega) + g_m(T^n\omega). \end{aligned}$$

Hence, $g_n(\omega)/n$ converges almost surely in $[-\infty,\infty)$, taking exponentials and noting that T was ergodic, the limit of (4.4) is non-negative and constant almost surely.

References

[AN72] K. B. Athreya and P. E. Ney. *Branching processes.* Springer-Verlag, New York-Heidelberg, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[Ath94] K. B. Athreya. Large deviation rates for branching processes. I. Single type case. *Ann. Appl. Probab.*, **4**, no. 3, (1994), 779–790.

[Fra14] J. M. Fraser. Assouad type dimensions and homogeneity of fractals. *Trans. Amer. Math. Soc.*, **366**, (2014), 6687–6733.

[GH17] I. García and K. Hare. Properties of Quasi-Assouad dimension. *ArXiv e-prints.*

[GW75] F. Galton and H. W. Watson. On the probability of the extinction of families. *Journal of the Royal Anthropological Institute*, **4**, (1875), 138–144.

[Kin73] J. F. C. Kingman. Subadditive ergodic theory. *Ann. Probability*, **1**, (1973), 883–909. With discussion by D. L. Burkholder, Daryl Daley, H. Kesten, P. Ney, Frank Spitzer and J. M. Hammersley, and a reply by the author.

[Ste89] J. M. Steele. Kingman's subadditive ergodic theorem. *Ann. Inst. H. Poincaré Probab. Statist.*, **25**, no. 1, (1989), 93–98.

[Tro19] S. Troscheit. The quasi-Assouad dimension for stochastically self-similar sets (FirstView). *Royal Society of Edinburgh Proceedings A*, pp. 1–15.

[Wil91] D. Williams. *Probability with martingales.* Cambridge University Press, Cambridge, 1991.