Branching processes

Branching processes are often used to model population growth due to reproduction. Consider a sequence of random variables \((Z_0, Z_1, \ldots)\), where \(Z_n\) may be viewed as the number of individuals in the \(n\)-th generation. See the figure given below.

![Branching processes diagram](image)

1.1 Galton-Watson branching process

Let us begin with the simplest example of branching process known as Galton-Watson branching process. It is built up on the following assumptions:

i. \(Z_0 = 1\);

ii. each individual lives for one unit time and then on death produces \(k\) many offsprings with probability \(p_k\) (where we have \(\sum k p_k = 1\)).

iii. all offsprings behave independently.

If we denote by \(X_{n,i}\) the number of offsprings produced by the \(i\)-th member of generation \(n\), then we have

\[
Z_n = X_{n,1} + X_{n,2} + \ldots + X_{n,Z_n}.
\]

The assumptions made in Galton-Watson model as mentioned above would imply that \(X_{n,j}\) s are independent.

Theorem 1.1.1 (Mean and variance of population size). Let \(E[Z_1] = \mu < \infty\), and let \(\text{Var}(Z_1) = \sigma^2 < \infty\). Then,

\[
* \ E[Z_n] = \mu^n;
\]

\[
* \ \text{Var}(Z_n) = \sigma^2 \mu^{n-1}(1 + \mu + \ldots + \mu^{n-1})
\]
Proof. First of all we notice that,

\[ \mathbf{E}[Z_n] = \mathbf{E}[\mathbf{E}[Z_n|Z_{n-1}]] \]
\[ = \mathbf{E}[\mu \cdot Z_{n-1}] \]
\[ = \mu \cdot \mathbf{E}[Z_{n-1}] \]
\[ = \mu^n \]

Now we notice that,

\[ \mathbf{E}[(Z_n - \mu Z_{n-1})^2] = \mathbf{E}[\mathbf{E}[(Z_n - \mu Z_{n-1})^2|Z_{n-1}]] \]
\[ = \mathbf{E}(\text{var}(Z_n|Z_{n-1})) = \mathbf{E}[\sigma^2 Z_{n-1}] \]
\[ = \sigma^2 \cdot \mu^{n-1} \]

Thus, \( \mathbf{E}[Z_n^2] - 2\mu \mathbf{E}[Z_n Z_{n-1}] + \mu^2 \mathbf{E}[Z_{n-1}]^2 = \sigma^2 \cdot \mu^{n-1} \). Next we notice that,

\[ \mathbf{E}[Z_n Z_{n-1}] = \mathbf{E}[\mathbf{E}[Z_n Z_{n-1}|Z_{n-1}]] \]
\[ = \mathbf{E}[Z_{n-1} \mathbf{E}[Z_n|Z_{n-1}]] = \mathbf{E}[Z_{n-1} \mu Z_{n-1}] \]
\[ = \mu \mathbf{E}[Z_{n-1}^2]. \]

Thus we get

\[ \mathbf{E}[Z_n^2] = \sigma^2 \mu^{n-1} + \mu^2 \mathbf{E}[Z_{n-1}^2], \]

and

\[ \text{var}(Z_n) = \mathbf{E}[Z_n^2] - \mathbf{E}[Z_n]^2 \]
\[ = \mu^2 \mathbf{E}[Z_{n-1}^2] + \sigma^2 \mu^{n-1} - \mu^2 \mathbf{E}[Z_{n-1}]^2 \]
\[ = \mu^2 \text{var}(Z_{n-1}) + \sigma^2 \mu^{n-1} \]
\[ = \mu^4 \text{var}(Z_{n-2}) + \sigma^2 (\mu^{n-1} + \mu^n) \]
\[ = \ldots \]
\[ = \mu^{2(n-1)} \text{var}(Z_1) + \sigma^2 (\mu^{n-1} + \ldots + \mu^{2n-3}) \]
\[ = \sigma^2 \cdot \mu^{n-1} \cdot (1 + \mu + \ldots + \mu^{n-1}). \]

This completes the proof. \( \square \)
1.1.1 Probability of extinction

**Extinction** refers to the event that the population eventually dies out. More precisely, it refers to the event $\bigcup_{n=1}^{\infty}(Z_n = 0)$. We can speak about the probability of this event, which we usually refer to as the *extinction probability*.

As a consequence of a previous theorem, notice that if $\mu < 1$, then we shall have

$$\mathbb{E} \left[ \sum_{n=1}^{\infty} Z_n \right] = \frac{1}{1-\mu} < \infty,$$

and thus $P(\sum_n Z_n < \infty) = 1$, and hence the probability of extinction is $= 1$.

**Corollary 1.1.2.** If $\mu < 1$, then the probability of extinction $= 1$.

*Proof.* Let $\tau_e := \min\{n : Z_n = 0\}$ be the extinction time. Then, it is easy to notice that the extinction probability $p_e = P\{\tau_e < \infty\}$.

Now, notice that the event $\{\tau_e > n\}$ coincides with the event $\{Z_n \geq 1\}$.

Hence, by Markov’s inequality, it would follow that

$$P\{\tau_e > n\} = P\{Z_n \geq 1\} \leq \mathbb{E}[Z_n] = \mu^n.$$

Finally, it remains to notice that as $n \to \infty$, the value of $\mu^n \to 0$, and so the probability $P\{\tau_e > n\} \to 0$, so that $P\{\tau_e < \infty\} = 1$.

This completes the proof. \(\square\)

**Exercise 1.1.1.** Show that if $\mu < 1$, and $\sigma^2 < \infty$, then $Z_n$ should converge to $0$ in $L_2$.

Now, notice that the branching chain exhibits the property of *self-similarity* in the sense that if we consider any individual of any generation, then in terms of producing offsprings it behaves as the individual in the 0-th generation from whom the entire branching process started. This in fact resembles a more well known property, namely the Markov property. In fact, it is easy to notice that the sequence $(Z_0, Z_1, \ldots)$ is indeed a Markov chain with initial distribution of $Z_0$ given by $\delta_1$. Viewing 0 as a state of this Markov chain, we can notice that 0 is actually an *absorbing state* of the chain; once you have $Z_n = 0$, then $Z_{n+\ell} = 0$, $\forall \ell > 0$.

**Proposition 1.1.3.** If $p_i \neq 1$, then all non zero states of the branching process (when viewed as a Markov chain) are transient.

*Proof.* All we need to show is that for every non-zero state $i$, the following holds.

$$\rho_i := P\{Z_n = i \text{ for some } n \in \mathbb{N} \mid Z_0 = i\} < 1.$$ 

Consider two cases :

- **Case 1:** $p_0 > 0$.
  For this case notice that, for any non zero state $i$, we have

  $$\rho_i \leq P\{Z_1 > 0 \mid Z_0 = i\} = 1 - P\{Z_1 = 0 \mid Z_0 = i\} = 1 - p_0^i < 1.$$
Case 2: $p_0 = 0$.

For this case notice that, for any non zero state $i$, we must have

$$
\rho_i \leq P\{Z_1 = i | Z_0 = i\} = p_1^i < 1.
$$

Combining the conclusions drawn from the above casework, it follows that for all non zero states $i$, the following is true:

$$
P\{Z_n = i \text{ for some } n \in \mathbb{N} | Z_0 = i\} < 1.
$$

This completes the proof.

As a consequence of the above proposition, we get the following:

**Corollary 1.1.4.** If $p_1 < 1$, then $P\{Z_n \to 0, \text{ or } Z_n \to \infty\} = 1$.

**Sketch of proof.** Use the previous transitivity of non zero states, with the readily observable fact that a bounded sequence of non-negative integers must take some value infinitely often (thus any such non zero state must be visited infinitely often).

**Exercise 1.1.2.** What happens when $p_1 = 1$?

Before proceeding further let us state a result the proof of which relies on an easy usage of the infamous Borel-Cantelli lemma and is left as an exercise for the reader.

**Theorem 1.1.5.** If $(X_1, X_2, \ldots \ldots)$ is a sequence of random variables and $X$ is a random variable, such that for all $\varepsilon > 0$, we have

$$
\sum_{n=1}^{\infty} P\{|X_n - X| > \varepsilon\} < \infty,
$$

then we have $X_n$ converges almost surely to $X$.

**Exercise 1.1.3.** Prove Theorem 1.1.5.

Using this result we get the following:

**Corollary 1.1.6.** If $E[Z_1] = \mu < 1$, then $Z_n$ converges almost surely to 0.

**Proof.** Let $\varepsilon > 0$. Then, it is easy to notice that $P\{Z_n \geq \varepsilon\} \leq P\{Z_n \geq 1\}$ for any $n \geq 0$. Now, by Markov’s inequality, we get $P\{Z_n \geq 1\} \leq E[Z_n] = \mu^n$.

Thus for any $\varepsilon > 0$, we would have:

$$
\sum_{n=0}^{\infty} P\{|Z_n - 0| \geq \varepsilon\} \leq \sum_{n=0}^{\infty} \mu^n < \infty,
$$

for $\mu < 1$.

Hence by an application of the above theorem, we get $Z_n$ converges almost surely to 0.

This completes the proof.

**Exercise 1.1.4.** Simulate a branching chain where the offspring distribution $\sim$ Poisson distribution with some mean $\lambda$. Varying $\lambda$ how the dynamics of the branching process changes. Try to predict how they depend on $\lambda$.

The regime $\mu < 1$ is called the **subcritical regime**. And those of $\mu = 1$ and $\mu > 1$ are respectively called the **critical** and **supercritical** regimes.
1.1.2 Generating function for Galton-Watson branching process

We begin by recalling the definition of probability generating functions.

**Definition 1.1.1.** For a discrete random variable $X$, having probability mass function (pmf) $(f_k)_{k \geq 0}$, the probability generating function (pgf) of $X$ is the function $G_X$ defined by

$$G_X(z) = \mathbb{E}[z^X] = \sum_{k \geq 0} f_k z^k.$$  

Notice that, as $\sum_{k \geq 0} f_k = 1$, so $G_X$ is well defined for $|z| \leq 1$. Thus if we view $G_X$ as a power series, then it has radius of convergence $\geq 1$. It can also be noticed that $G_X$ is continuous for $|z| < 1$. And in fact an application of Abel’s theorem for power series would imply that $G_X$ is also continuous at the point 1. Thus we get the following :

**Proposition 1.1.7.** $G_X$ is continuous over $[0, 1]$.

*Proof.* Follows from the above discussions. \hfill $\square$

Using conditional expectations, and the definition of probability generating functions, we get the following.

**Lemma 1.1.8.** If $X$ is some discrete random variable, $X_1, X_2, \ldots$ are independent and identically distributed copies of $X$, and if $N$ is a discrete non-negative random variables independent of the $X_i$s, then the pgf of the random sum $S_N = \sum_{i=1}^N X_i$, can be given by

$$G_{S_N}(z) = G_N(G_X(z)).$$

*Proof.* Using properties of conditional expectation, we get

$$\mathbb{E}[z^{S_N}] = \mathbb{E} \mathbb{E}[z^{X_1 + \ldots + X_N} | N]$$

$$= \mathbb{E}_N \mathbb{E}[z^{X_1}] \cdot \ldots \cdot \mathbb{E}[z^{X_N}]$$

$$= \mathbb{E}_N [(G_X(z))^N] = G_N(G_X(z)).$$

This completes the proof. \hfill $\square$

As a consequence of the above lemma in the setting of Galton-Watson branching process, the relation $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$, would yield

**Corollary 1.1.9.** For every $n \geq 0$, we have

$$G_{Z_{n+1}}(z) = G_{Z_n}(G_X(z)).$$

As a consequence,

$$G_{Z_n}(z) = G_X \circ G_X \circ \ldots \circ G_X(z),$$

where on the right hand side $G_X$ is composed $n$ many times.

As yet another consequence of the above corollary, we get the following:
Corollary 1.1.10. The probability $p_{e,n}$ of the event of extinction by time $n$, can be expressed as

$$p_{e,n} = G_X \circ G_X \circ \ldots \circ G_X(0).$$

The following proposition guarantees a solution to the equation $z = G_X(z)$ in the interval $[0, 1]$.

Proposition 1.1.11. For a Galton-Watson branching process $G_X(1) = 1$, where $G_X$ denotes the pgf of the offspring distribution.

Proof. This directly follows from the definition of a probability generating function. □

Now, we state and prove an interesting result as follows:

Lemma 1.1.12. The extinction probability $p_e$ is the smallest solution to the equation $z = G_X(z)$ in the interval $[0, 1]$.

Proof. First of all we notice that $p_e$ must satisfy $p_e = G_X(p_e)$. To show this, we notice that $G_X$ is continuous. Now, we observe that

$$p_e = \lim_{n \to 0} P(Z_n = 0).$$

Then since, $P(Z_n = 0) = G_{Z_n}(0)$, so

$$G_X(p_e) = G_X\left(\lim_{n \to \infty} G_{Z_n}(0)\right) = \lim_{n \to \infty} G_X(G_{Z_n}(0)) = \lim_{n \to \infty} G_{Z_n+1}(0) = p_e.
\quad □$$

The following exercise discusses an important result in the theory of Galton Watson branching processes.

Exercise 1.1.5. For a non trivial Galton Watson branching process, if $p_1 \neq 1$, and $\mu \leq 1$, then prove that the probability of extinction $p_e$ is equal to 1. Show that on the other hand any other non trivial Galton Watson branching process must have its extinction probability $p_e < 1$.

The following result gives a very powerful tool to study certain characteristics of Galton-Watson branching processes.

Proposition 1.1.13. $\sup_{t \in [0, p_e]} |p_e - G_{Z_n}(t)| \to 0$ as $n \to \infty$.

Proof. For each $t \in [0, p_e]$, and $n \in \mathbb{Z}_{\geq 0}$, we define $f_n(t) = G_{Z_n}(t)$.

Then, notice that, all we need to show is that the sequence $(f_n)$ of functions converges uniformly to the constant function $p_e$.

First we notice that, these functions $(f_n)$ are montonically increasing (i.e., $f_{n+1}(t) = G_X(f_n(t)) \geq f_n(t)$, for any $n \in \mathbb{Z}_{\geq 0}$), continuous and are defined on a compact set. So, if we can show that $(f_n)$ converges pointwise to $p_e$ then by invoking Dini’s theorem, we can conclude that $(f_n)$ also converges uniformly to $p_e$.

So, it suffices to show that $(f_n)$ converges pointwise to $p_e$.

To show this, we fix an arbitrary $t \in [0, p_e]$. If $t \in \{0, p_e\}$, then it is obvious, since then $G_X(t) = p_e$.

So, now let $t \in (0, p_e)$. Then, notice that since $p_e$ is the smallest solution of $G_X(z) = z$ in
the interval $[0, 1]$, so $f_{n+1}(t) \geq f_n(t)$ for every positive integer $n$, and $f_n(t) \leq p_e$ for every positive integer $n$. And hence by Bolzano Weirestrass theorem, it follows that $f_n(t)$ must converge to some real number $r_t$ in $[0, 1]$. Now, notice that since $p_e$ is the smallest solution to $G_X(z) = z$ in $[0, p_e]$, so immediately it follows that $f_n(t)$ converges to $p_e$, for each real number $t$. This proves the pointwise convergence, and since the constant function $p_e$ is continuous, so invoking Dini’s theorem, yields the uniform convergence we wanted to establish.

This completes the proof.

Before we proceed further, let us see some obvious results related to the extinction probability.

- First of all notice that if $p_0 = 0$, then obviously the process cannot go extinct, and thus $p_0 = 0 \implies p_e = 0$.
- If $p_0 > 0$, and $p_0 + p_1 = 1$, then obviously the probability that the process doesn’t go extinct is the probability that at each step a single offspring is produced from the individuals of the current generation and in this case this probability will be $= 0$, since the probability that each of first $n$ consecutive generations give birth of exactly one offspring $= p_0^n \to 0$ as $n \to \infty$.

To make things more interesting let us keep aside these trivial cases, and assume $p_0 > 0$, and $p_0 + p_1 < 1$. We call such a Galton Watson branching process **non trivial**.

For $n \in \mathbb{N}$, we will write $W_n$ to denote the random variable $\frac{Z_n}{\mu^n}$.

**Proposition 1.1.14.** If $p_1 \neq 1$, then $W_n \to 0$ almost surely in the subcritical and critical regimes.

**Proof.** First let us address the case of subcritical regime.

To begin with notice that $(\mu < 1 \implies \sum_{n \geq 1} \mathbb{E}[Z_n] < \infty$ holds in the subcritical regime. Now, in view of theorem 1.1.5, it suffices to show that for all $\varepsilon > 0$, we have

$$
\sum_{n \geq 1} P\{Z_n > \mu^n \cdot \varepsilon\} < \infty.
$$

To show this, we notice that the event $\{Z_n > \mu^n \cdot \varepsilon\}$ is a subset of the event $\{Z_n > 0\}$ and hence

$$
\sum_{n \geq 1} P\{Z_n > \mu^n \cdot \varepsilon\} < \sum_{n \geq 1} P\{Z_n > 0\} = \sum_{n \geq 1} P\{Z_n \geq 1\}.
$$

Now, by Markov’s inequality, the right hand side of the above can be bounded by $\sum_{n \geq 1} \mathbb{E}[Z_n]$ which is finite (as deduced above).

Thus, we conclude that for the subcritical regime $W_n$ converges to 0 almost surely.

Now, let us turn our attention towards the critical regime.

Before we do anything notice that the condition $p_1 \neq 1$ is necessary ; for $p_1 = 1$, we will have $W_n = 1$ for all $n$.

Now suppose $p_1 \neq 1$. Then by a previous result we have

$$
P\{Z_n \to 0 \text{ or } Z_n \to \infty\} = 1.
$$
So, it suffices to show that \( P\{Z_n \to \infty\} = 0 \). Notice that in the critical regime (\( \mu = 1 \implies \)) \( Z_n = W_n \) for all \( n \).

So, equivalently it suffices to show that \( P\{W_n \to \infty\} = 0 \).

Notice that if \( P\{W_n \to \infty\} = q \neq 0 \), then for all \( M > 0 \), there exists some \( k = k(M) \in \mathbb{N} \), such that
\[
P\{\omega \in \Omega : W_k(\omega) > M\} \geq q/2.
\]

This is because, if for all \( k \in \mathbb{N} \), we have \( P\{\omega \in \Omega : W_k(\omega) > M\} = g_k < q/2 \), then \( \lim_{k \to \infty} g_k \leq q/2 < q \).

However, notice that \( \lim_{k \to \infty} g_k \) is actually the probability
\[
P\{\omega \in \Omega : W_k(\omega) > M \text{ for some } k \in \mathbb{N}\} \geq q.
\]

Now, notice that
\[
\mathbb{E}[W_n] = \sum_{m \geq 1} P\{W_n \geq m\}.
\]

Now, notice that if \( W_n \to \infty \) with a strictly positive probability say \( q \), then for a sufficiently large \( n \), we will have \( P\{W_n \geq m\} \geq q/2 > 0 \), for all \( m \in \mathbb{N} \cap [1, 2+5/q] \). Hence,
\[
\mathbb{E}[W_n] = \sum_{m \geq 1} P\{W_n \geq m\} \geq \sum_{m \in \mathbb{N} \cap [1, 2+5/q]} q > 1.
\]

This violates the fact that \( \mathbb{E}[W_n] = 1 \).

Thus, we cannot have \( P\{Z_n \to \infty\} \neq 0 \). Hence in conclusion, \( P\{Z_n \to 0\} = 1 \), for the critical regime for \( p_1 \neq 1 \).

In other words, for the critical regime we must have \( p_1 \neq 1 \implies W_n \to 0 \) almost surely.

\[\diamondsuit\]

This completes the proof.

\[\square\]

Using Lemma 2.2.6 one can prove the following useful theorem. Then, we can give yet another proof of Proposition 1.1.14 using Exercise 1.1.5.

**Corollary 1.1.15.** Proposition 1.1.14 is true.

**Proof.** The proof for the subcritical regime can be done exactly how we proved Proposition 1.1.14 above.

For the critical regime it suffices to notice that \( W_n = Z_n \) for all positive integers \( n \), and Theorem 2.2.8 implies that \( Z_n \) converges to 0 almost surely, and hence so does \( W_n \).

\[\square\]

Yet another proof of Proposition 1.1.14 can be given using Doob’s martingale convergence theorem, which we won’t discuss here, since it demands more prerequisites.

Before we proceed further, let us see some examples, where we can find an expression for the extinction probabilities.
**Example 1.1.16.** We know that the Poisson ($\lambda$) distribution has the pgf $s \mapsto e^{-\lambda(1-s)}$. Thus, if we have a branching process whose offspring distribution is Poisson ($\lambda$), then the extinction probability is the smallest solution to the equation

$$s = e^{-\lambda(1-s)}$$

in the interval $[0,1]$. Hence, in this case the extinction probability $p_e$ equals the smallest solution to the equation

$$\log s - \lambda s = -\lambda.$$

It can be found numerically.

**Example 1.1.17.** Suppose we have a branching process where the offspring distribution has law geometric ($p$). Then, we know that the pgf of geometric ($p$) distribution is given by $s \mapsto \frac{1-p}{1-ps}$. Thus, in this case the extinction probability will be the smallest solution to the equation $s = \frac{1-p}{1-ps}$ in the interval $[0,1]$. This means that the extinction probability $p_e$ is the smallest solution to the equation

$$s^2 p - s + 1 - p = 0$$

in the interval $[0,1]$. This is a quadratic equation, and hence we can obtain its roots using the infamous Sreedharacharya’s formula. Then, it suffices to look for the smallest root in $[0,1]$.

In statistics, a **zero-inflated model** is a statistical model based on a zero-inflated probability distribution, i.e. a distribution that allows for frequent zero-valued observations. One such example is the **zero-inflated geometric distribution** described as follows:

We say that a random variable $X$ has law zero-inflated geometric with parameters $b,c$ if

$$P(X = k) = \begin{cases} b^{k-1} \cdot c, & \text{if } k \geq 1, \\ P(X = 0) & \text{is adjusted accordingly so that the sum of probabilities } = 1. \end{cases}$$

**Example 1.1.18.** Suppose we have a branching process whose offspring distribution is zero-inflated geometric with parameters $b,c$.

Notice that if $X \sim$ zero-inflated geometric with parameters $b,c$, then

$$P(X = 0) = 1 - \sum_{k \geq 1} b^{k-1} \cdot c = 1 - c \cdot \sum_{g \geq 0} b^g = 1 - \frac{c}{1 - b}.$$

From here, it is easy to see that the pgf of this distribution is given by:

$$s \mapsto \frac{c}{b} \cdot \left( \frac{1}{1 - bs} - 1 \right) + 1 - \frac{c}{1 - b}.$$

Thus, the extinction probability can be given by the smallest solution to the equation

$$s = \frac{c}{b} \cdot \left( \frac{1}{1 - bs} - 1 \right) + 1 - \frac{c}{1 - b}.$$

This can be modified to obtain

$$s^2(b^3 - b^2) + s(2b - b^2 - bc) + cb - 1 = 0.$$

This is a quadratic equation, and hence we can obtain its roots using the infamous Sreedharacharya’s formula. Then, it suffices to look for the smallest root in $[0,1]$. 

14
Another useful result is recorded below.

**Theorem 1.1.19.** If $\mu > 1$, and $\sigma^2 < \infty$, then $W_n$ converges almost surely as well as in $L_2$ to some non deterministic random variable $W_\infty$ having mean 1 and variance $\frac{\sigma^2}{\mu(\mu-1)}$.

Before we present the proof, let us collect some ingredients which we shall require to prove the grand Theorem 1.1.19. We begin with an elementary observation, proving which rigorously is left as an exercise.

**Exercise 1.1.6.** Show that the sequence $(W_n)$ of random variables satisfies the Markov property.

Next, we turn to utilising this Markov property.

**Proposition 1.1.20.** $\mathbb{E}[W_{n+k} | W_n, \ldots, W_0] = \mathbb{E}[W_{n+k} | W_n] = W_n$.

**Proof.** The first equality follows directly from the Markov property of $(W_n)$.

Now, we notice that,

$$\mathbb{E}[W_{n+k} | Z_n] = \mu^k Z_n = \mu^{n+k} W_n.$$ 

Thus, we conclude that

$$\mathbb{E}[W_{n+k} | W_n] = \frac{1}{\mu^{n+k}} \cdot \mu^{n+k} \cdot W_n = W_n.$$ 

This completes the proof. □

The property of $(W_n)$ stated and subsequently proved in Proposition 1.1.20, is popularly known as the *martingale* property.

**Proposition 1.1.21.** We have the following:

$$\mathbb{E}[(W_{n+k} - W_n)^2] = \frac{\sigma^2 \mu^{-n} (1 - \mu^{-k})}{\mu(\mu-1)}.$$ 

**Proof.** To begin with, we notice that from Proposition 1.1.20, it follows that

$$\mathbb{E}[W_{n+k} W_n | W_n] = W_n \cdot \mathbb{E}[W_{n+k} | W_n] = W_n^2.$$ 

Hence by taking another expectation, and using the tower property of double expectation, we get

$$\mathbb{E}[W_{n+k} W_n] = \mathbb{E}[W_n^2].$$ 

Thus, upon expanding $(W_{n+k} - W_n)^2$, we would get

$$\mathbb{E}[(W_{n+k} - W_n)^2] = \mathbb{E}[W_n^2] - 2\mathbb{E}[W_{n+k} W_n] + \mathbb{E}[W_n^2] = \mathbb{E}[W_n^2] - \mathbb{E}[W_n^2].$$ 

Now, we can find $\mathbb{E}[W_n^2]$ by writing

$$\mathbb{E}[W_n^2] = \text{var}(W_n) + (\mathbb{E}[W_n])^2 = \text{var}(W_n) + 1,$$
and then using the fact that $W_m = Z_m / \mu^m$, and Theorem 1.1.1, we can obtain
\[
\text{var}(W_m) = \sigma^2 \mu^{n-1}(1 + \ldots + \mu^{n-1}) \cdot \frac{1}{\mu^{2m}}.
\]
From this it follows that
\[
\mathbb{E}[(W_{n+k} - W_n)^2] = \frac{\sigma^2 \mu^{-n}(1 - \mu^{-k})}{\mu(\mu - 1)}.
\]
This completes the proof. \qed

We will also use the following result without proof (the proof can be done using Minkowski’s inequality and monotone convergence theorem and is classical).

**Theorem 1.1.22.** The space of all random variables with finite variance is complete with respect to the $L_2$ norm.

**Proof of Theorem 1.1.19.** To begin with, we notice that the Markov property of $(Z_n)$ implies the Markov property of $(W_n)$. Hence,
\[
\mathbb{E}[W_{n+k}|W_n, \ldots, W_0] = \mathbb{E}[W_{n+k}|W_n] = W_n.
\]
Next, we notice that using the above we can obtain
\[
\mathbb{E}[W_{n+k} - W_n]^2 = \frac{\sigma^2 \mu^{-n}(1 - \mu^{-k})}{\mu(\mu - 1)}.
\]
This implies $(W_n)$ is a Cauchy sequence with respect to the $L_2$ norm. Hence, invoking the result in Theorem 1.1.22, it follows that $(W_n)$ converges with respect to the $L_2$ norm to some random variable $W_\infty$.

Now, we notice that since $(W_n)$ converges to $W_\infty$ in $L_2$ norm, so the convergence also holds in $L_1$, and so $\mathbb{E}[|W_\infty|] = \lim_{n \to \infty} \mathbb{E}[|W_n|] = 1$. However, also notice that for each $n$, the random variable $W_n$ is a non-negative valued, and so is $W_\infty$. Hence, it follows that
\[
\mathbb{E}[W_\infty] = \mathbb{E}[|W_\infty|] = 1.
\]
Now let’s try to find the variance of $W_\infty$. Notice that,
\[
\text{var}(W_\infty) = \mathbb{E}[W_\infty^2] - (\mathbb{E}[W_\infty])^2 = \mathbb{E}[W_\infty^2] - 1.
\]
Now, notice that since $(W_n)$ converges to $W_\infty$ in $L_2$, so $\mathbb{E}[W_\infty^2] = \lim_{n \to \infty} \mathbb{E}[W_n^2]$.

Now, notice that for any $n$, we have $\mathbb{E}[W_n^2] = \text{var}(W_n) + \mathbb{E}[W_n]^2 = \text{var}(W_n) + 1$. Also notice that, $W_n = Z_n / \mu^n \implies \text{var}(W_n) = \text{var}(Z_n) / \mu^{2n}$. Now, as computed earlier, we know that $\text{var}(Z_n) = \sigma^2 \mu^{n-1}(1 + \ldots + \mu^{n-1})$. Thus it follows that,
\[
\mathbb{E}[W_\infty^2] = \lim_{n \to \infty} \left(\sigma^2 \mu^{n-1}(1 + \ldots + \mu^{n-1}) + 1\right) \cdot \frac{1}{\mu^{2n}}.
\]
This implies that,
\[
\text{var}(W_\infty) = \frac{\sigma^2}{\mu \cdot (\mu - 1)}.
\]
The almost sure convergence follows from the fact that

$$\mathbb{E} \left[ \sum_{n \geq 1} (W_n - W_{\infty})^2 \right] = \sum_{n \geq 1} \mathbb{E}[(W_n - W_{\infty})^2] < \infty,$$

where Tonelli’s theorem allows us to rearrange the series with impunity, to obtain the first equality.

Combining our deductions, we conclude that \((W_n)\) converges almost surely as well as in \(L_2\) to some random variable \(W_{\infty}\) having mean 1, and variance \(\frac{\sigma^2}{\mu(\mu-1)}\).

This completes the proof. \(\square\)

Now, that we know the mean and variance of the limiting random variable \(W_{\infty}\), we may be interested in finding its distribution. Let us try to figure out the moment generating function (i.e., \(t \mapsto \mathbb{E}[e^{tW_{\infty}}]\)) of the random variable \(W_{\infty}\).

Let us proceed in steps as follows:

- Notice that \(W_n\) converges almost surely to \(W_{\infty}\) implies that \(W_n\) converges in distribution to \(W_{\infty}\).
- Hence by Portmanteau lemma, it follows that \(\mathbb{E}[e^{tW_n}]\) converges to \(\mathbb{E}[e^{tW_{\infty}}]\).
- For any \(n \in \mathbb{N}\), notice that \(\mathbb{E}[e^{tW_n}] = \sum_{z \geq 0} e^{tz/\mu^n} \cdot P\{Z_n = z\}\). It is easy to see that the sum on the RHS of this equation converges whenever \(t \leq 0\). Also it shows us that the moment generating function of \(W_{\infty}\) which we are looking for is the limit of the quantity \(\mathbb{E}[e^{tW_n}]\) as \(n \to \infty\).
- Thus, we see that \(\mathbb{E}[e^{tW_{\infty}}] = \lim_{n \to \infty} \sum_{z \geq 0} e^{tz/\mu^n} \cdot P\{Z_n = z\}\).

Let us see if we can tell more about the mgf.

Consider the mgf \(\phi_{W_n}\) of \(W_n\). Then notice that,

\[
\phi_{W_n}(\mu t) = \mathbb{E}[e^{\mu t W_n}]
= G_{Z_n}(e^{t/\mu^{n-1}})
= G_{X}(G_{Z_{n-1}}(e^{t/\mu^{n-1}}))
= G_{X}(\mathbb{E}[e^{Z_{n-1}t/\mu^{n-1}}])
= G_{X}(\mathbb{E}[e^{W_{n-1}t}]) = G_{X}(\phi_{W_{n-1}}(t)).
\]

This alongwith a previously derived result on convergence of the mgf of \(W_n\) to that of \(W_{\infty}\) allows us to conclude that:

**Proposition 1.1.23.** If \(\phi_{W_{\infty}}\) denotes the moment generating function of \(W_{\infty}\), and \(G_{X}\) denotes the probability generating function of the offspring distribution, then \(\phi_{W_{\infty}}(t)\) is finite for all \(t \leq 0\), and we have the following:

\[
\phi_{W_{\infty}}(\mu t) = G_{X}(\phi_{W_{\infty}}(t)).
\]

**Proof.** Follows from the above discussions. \(\square\)
In some cases Proposition 1.1.23 helps us to study the moment generating function in some cases (for instance if $G_X$ is linear; see Exercise 1.1.7), and hence understand the distribution of $W_\infty$ better in these cases.

**Exercise 1.1.7.** Considering the case when $G_X$ is linear, prove Proposition 1.1.14 in this case, using Proposition 1.1.23.