1.1. Introduction

Branching processes are a fundamental object in probability theory. They serve as models for the reproduction of particles or individuals within a collective or a population. Here we act on the assumption that the population evolves within clearly distinguishable generations, which allows us to examine the population at the founding generation $n = 0$ and the subsequent generations $n = 1, 2, \ldots$. To begin with, we focus on the sequence of population sizes $Z_n$ at generation $n, n \geq 0$. Later, we shall study whole family trees.

Various kinds of randomness can be incorporated into such branching models. For this monograph, we have two such types in mind. On the one hand, we take randomness in reproduction into account. Here a main assumption is that different individuals give birth independently and that their offspring distributions coincide within each generation. On the other hand, we consider environmental stochasticity. This means that these offspring distributions may change at random from one generation to the next. A fundamental question concerns which one of the two random components will dominate and determine primarily the model’s long-term behavior. We shall get to know the considerable influence of environmental fluctuations.

This first chapter is of a preliminary nature. Here we look at branching models with reduced randomness. We allow that the offspring distributions vary among the generations but as a start in a deterministic fashion. So to speak we consider the above model conditioned by its environment.
We begin with introducing some notation. Let $\mathcal{P}(\mathbb{N}_0)$ be the space of all probability measures on the natural numbers $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. For $f \in \mathcal{P}(\mathbb{N}_0)$, we denote its weights by $f[z]$, $z = 0, 1, \ldots$. We also define

$$f(s) := \sum_{z=0}^{\infty} f[z] s^z, \quad 0 \leq s \leq 1.$$ 

The resulting function on the interval $[0, 1]$ is the generating function of the measure $f$. Thus, we take the liberty here to denote the measure and its generating function by *one and the same* symbol $f$. This is not just as probability measures and generating functions uniquely determine each other but operations on probability measures are often most conveniently expressed by means of their generating functions. Therefore, for two probability measures $f_1$ and $f_2$, the expressions $f_1 f_2$ or $f_1 \circ f_2$ do not only stand for the product or composition of their generating functions but also stand for the respective operations with the associated probability measures (in the first case, it is the convolution of $f_1$ and $f_2$). Similarly, the derivative $f'$ of the function $f$ may be considered as well as the measure with weights $f'[z] = (z + 1) f[z + 1]$ (which in general is no longer a probability measure). This slight abuse of notation will cause no confusions but on the contrary will facilitate presentation. Recall that the mean and the normalized second factorial moment,

$$\bar{f} := \sum_{z=1}^{\infty} z f[z] \quad \text{and} \quad \tilde{f} := \frac{1}{f'} \sum_{z=2}^{\infty} z(z-1) f[z]$$

can be obtained from the generating functions as

$$\bar{f} = f'(1), \quad \tilde{f} = \frac{f''(1)}{f'(1)^2}.$$ 

NOTE.– Any operation we shall apply to probability measures (and more generally to finite measures) on $\mathbb{N}_0$ has to be understood as an operation applied to their generating functions.

We are now ready for first notions. Let $(\Omega, \mathcal{F}, P)$ be the underlying probability space.

**Definition 1.1.**– A sequence $v = (f_1, f_2, \ldots)$ of probability measures on $\mathbb{N}_0$ is called a varying environment.

**Definition 1.2.**– Let $v = (f_n, n \geq 1)$ be a varying environment. Then a stochastic process $Z = \{Z_n, n \in \mathbb{N}_0\}$ with values in $\mathbb{N}_0$ is called a branching process with environment $v$, if for any integers $z \geq 0, n \geq 1$

$$P(Z_n = z \mid Z_0, \ldots, Z_{n-1}) = (f_n^z[z]) \quad P\text{-a.s.}$$
On the right-hand side, we have the $Z_{n-1}$th power of $f_n$. In particular, $Z_n = 0$ $\mathbb{P}$-a.s. on the event that $Z_{n-1} = 0$. If we want to emphasize that probabilities $\mathbb{P}(\cdot)$ are determined on the basis of the varying environment $v$, we use the notation $\mathbb{P}_v(\cdot)$.

In probabilistic terms, the definition says, for $n \geq 1$, that given $Z_0, \ldots Z_{n-1}$ the random variable $Z_n$ may be realized as the sum of i.i.d. random variables $Y_{i,n}$, $i = 1, \ldots, Z_{n-1}$, with distribution $f_n$,

$$Z_n = \sum_{i=1}^{Z_{n-1}} Y_{i,n}.$$

This corresponds to the following conception of the process $Z$: $Z_n$ is the number of individuals of some population in generation $n$, where all individuals reproduce independently of each other and of $Z_0$, and where $f_n$ is the distribution of the number $Y_n$ of offspring of an individual in generation $n - 1$. The distribution of $Z_0$, which is the initial distribution of the population, may be arbitrary. Mostly we choose it to be $Z_0 = 1$.

**EXAMPLE 1.1.–** A branching process with the constant environment $f = f_1 = f_2 = \cdots$ is called a Galton–Watson process with offspring distribution $f$. \hfill $\Box$

The distribution of $Z_n$ is conveniently expressed via composing generating functions. For probability measures $f_1, \ldots f_n$ on $\mathbb{N}_0$ and for natural numbers $0 \leq m < n$, we introduce the probability measures

$$f_{m,n} := f_{m+1} \circ \cdots \circ f_n. \tag{1.1}$$

Moreover, let $f_{n,n}$ be the Dirac measure $\delta_1$.

**PROPOSITION 1.1.–** Let $Z$ be a branching process with initial size $Z_0 = 1$ a.s. and varying environment $(f_n, n \geq 1)$. Then for $n \geq 0$, the distribution of $Z_n$ is equal to the measure $f_{0,n}$.

**PROOF.**– Induction on $n$. \hfill $\Box$

Usually it is not straightforward to evaluate $f_{0,n}$ explicitly. The following example contains an exceptional case of particular interest.

**EXAMPLE 1.2.–** LINEAR FRACTIONAL DISTRIBUTIONS. A probability measure $f$ on $\mathbb{N}_0$ is said to be of the linear fractional type, if there are real numbers $p, a$ with $0 < p < 1$ and $0 \leq a \leq 1$, such that

$$f[z] = apq^{z-1} \quad \text{for } z \neq 0,$$
with $q = 1 - p$. For $a > 0$, this implies

$$f[0] = 1 - a, \quad \bar{f} = \frac{a}{p}, \quad \tilde{f} = \frac{2q}{a}.$$

We shall see that it is convenient to use the parameters $\bar{f}$ and $\tilde{f}$ instead of $a$ and $p$. Special cases are, for $a = 1$, the geometric distribution $g$ with success probability $p$ and, for $a = 0$, the Dirac measure $\delta_0$ at point 0. In fact, $f$ is a mixture of both, i.e. $f = ag + (1 - a)\delta_0$. A random variable $Z$ with values in $\mathbb{N}_0$ has a linear fractional distribution, if

$$\mathbb{P}(Z = z \mid Z \geq 1) = pq^{z-1} \quad \text{for } z \geq 1,$$

that is, if its conditional distribution, given $Z \geq 1$, is geometric with success probability $p$. Then

$$\mathbb{P}(Z \geq 1) = a = \left( \frac{1}{\bar{f}} + \frac{\tilde{f}}{2} \right)^{-1}.$$

For the generating function, we find

$$f(s) = 1 - a \frac{1 - s}{1 -qs}, \quad 0 \leq s \leq 1$$

(leading to the naming of the linear fractional). It is convenient to convert it for $\bar{f} > 0$ into

$$\frac{1}{1 - f(s)} = \frac{1}{\bar{f} \cdot (1 - s)} + \frac{\tilde{f}}{2}, \quad 0 \leq s < 1. \quad [1.2]$$

Note that this identity uniquely characterizes the linear fractional measure $f$ with mean $\bar{f}$ and normalized second factorial moment $\tilde{f}$.

The last equation now allows us to determine the composition $f_{0,n}$ of linear fractional probability measures $f_k$ with parameters $\bar{f}_k, \tilde{f}_k$, $1 \leq k \leq n$. From $f_{0,n} = f_1 \circ f_{1,n}$,

$$\frac{1}{1 - f_{0,n}(s)} = \frac{1}{f_1 \cdot (1 - f_{1,n}(s))} + \frac{\tilde{f}_1}{2}.$$

Iterating this formula we obtain (with $\bar{f}_1 \cdots \bar{f}_{k-1} := 1$ for $k = 1$)

$$\frac{1}{1 - f_{0,n}(s)} = \frac{1}{f_1 \cdots f_n \cdot (1 - s)} + \frac{1}{2} \sum_{k=1}^n \frac{\tilde{f}_k}{f_1 \cdots f_{k-1}}. \quad [1.3]$$
It implies that the measure \( f_{0,n} \) itself is of the linear fractional type with a mean and normalized second factorial moment

\[
\tilde{f}_{0,n} = \frac{\sum_{k=1}^{n} \tilde{f}_k}{\tilde{f}_1 \cdots \tilde{f}_{k-1}}.
\]

This property of perpetuation is specific for probability measures of the linear fractional type. □

For further investigations, we now rule out some cases of less significance.

**Assumption V1.** – The varying environment \((f_1, f_2, \ldots)\) fulfills \(0 < \bar{f}_n < \infty\) for all \(n \geq 1\).

Note that, in the case of \(\bar{f}_n = 0\), the population will a.s. be completely wiped out in generation \(n\).

From Proposition 1.1, we obtain formulas for moments of \(Z_n\) in a standard manner. Taking derivatives by means of Leibniz’s rule and induction, we have, for \(0 \leq m < n\),

\[
f'_{m,n}(s) = \prod_{k=m+1}^{n} f'_k(f_k,n(s)),
\]

and \(f'_{n,n}(s) = 1\). In addition, using the product rule, we obtain after some rearrangements

\[
f''_{m,n}(s) = f'_{m,n}(s)^2 \sum_{k=m+1}^{n} \frac{f''_k(f_k,n(s))}{f'_k(f_k,n(s))^2 \prod_{j=m+1}^{k-1} f'_j(f_j,n(s))}, \quad [1.4]
\]

and \(f''_{n,n}(s) = 0\). Evaluating these equations for \(m = 0\) and \(s = 1\), we get the following formulas for means and normalized second factorial moments of \(Z_n\), which we had already come across in the case of linear fractional distributions (now the second factorial moments may well take the value \(\infty\)).

**Proposition 1.2.** – For a branching process \(Z\) with initial size \(Z_0 = 1\) a.s. and environment \((f_1, f_2, \ldots)\) fulfilling V1, we have

\[
\mathbb{E}[Z_n] = \tilde{f}_1 \cdots \tilde{f}_n, \quad \frac{\mathbb{E}[Z_n(Z_n - 1)]}{\mathbb{E}[Z_n]^2} = \sum_{k=1}^{n} \frac{\tilde{f}_k}{\tilde{f}_1 \cdots \tilde{f}_{k-1}}. \quad [1.5]
\]
We note that these equations entail the similarly built formula

\[
\frac{\text{Var}[Z_n]}{E[Z_n]^2} = \sum_{k=1}^{n} \frac{\rho_k}{f_1 \cdots f_{k-1}},
\]

set up for the standardized variances

\[
\rho_k := \frac{1}{f_k^2} \sum_{z=0}^{\infty} (z - \bar{f}_k)^2 f_k(z)
\]

of the probability measures \(f_k\). Indeed,

\[
\sum_{k=1}^{n} \frac{\rho_k}{f_1 \cdots f_{k-1}} = \sum_{k=1}^{n} \frac{\bar{f}_k''(1) + f_k'(1) - f_k'(1)^2}{f_1 \cdots f_{k-1} \cdot f_k^2}
\]

\[
= \sum_{k=1}^{n} \frac{\bar{f}_k''(1)}{f_1 \cdots f_{k-1} \cdot f_k^2} + \sum_{k=1}^{n} \left( \frac{1}{f_1 \cdots f_{k-1}} - \frac{1}{f_1 \cdots f_{k-1}} \right)
\]

\[
= \frac{E[Z_n(Z_n - 1)]}{E[Z_n]^2} + \frac{1}{E[Z_n]} - 1 = \frac{\text{Var}[Z_n]}{E[Z_n]^2}.
\]

1.2. Extinction probabilities

For a branching process \(Z\), let

\[
\theta := \min \{ n \geq 1 : Z_n = 0 \}
\]

be the moment when the population dies out. Then \(P(\theta \leq n) = P(Z_n = 0)\), and the probability that the population becomes ultimately extinct is equal to

\[
q := P(\theta < \infty) = \lim_{n \to \infty} P(Z_n = 0).
\]

In this section, we would like to characterize a.s. extinction. For a first criterion, we use the Markov inequality

\[
P(\theta > n) = P(Z_n \geq 1) \leq E[Z_n]
\]

and the fact that \(P(Z_n \geq 1)\) is decreasing in \(n\). We obtain

\[
\lim_{n \to \infty} \bar{f}_1 \cdots \bar{f}_n = 0 \quad \Rightarrow \quad q = 1.
\]
On the other hand, the Paley–Zygmund inequality tells us that
\[
P(\theta > n) = P(Z_n > 0) \geq \frac{\mathbb{E}[Z_n]^2}{\mathbb{E}[Z_n^2]} = \frac{\mathbb{E}[Z_n]^2}{\mathbb{E}[Z_n] + \mathbb{E}[Z_n(Z_n - 1)]},
\]
which in combination with equation [1.6] yields the bound
\[
\frac{1}{P(\theta > n)} \leq \frac{\mathbb{E}[Z_n^2]}{\mathbb{E}[Z_n]^2} = 1 + \frac{\text{Var}[Z_n]}{\mathbb{E}[Z_n]^2} = 1 + \sum_{k=1}^{n} \frac{\rho_k}{f_1 \cdots f_{k-1}}.
\]

Thus, the question arises as to which one of both bounds captures the size of \(P(\theta > n)\) more adequately. It turns out that, under a mild extra assumption, it is the Paley–Zygmund bound.

**Assumption V2.** For the varying environment \((f_1, f_2, \ldots)\), there exists a constant \(c < \infty\) such that for all \(n \geq 1\)
\[
\mathbb{E}[Y_n(Y_n - 1)] \leq c \mathbb{E}[Y_n] \cdot \mathbb{E}[Y_n - 1 | Y_n > 0],
\]
where the random variables \(Y_1, Y_2, \ldots\) have the distributions \(f_1, f_2, \ldots\)

This uniformity assumption is typically satisfied, as illustrated by the following examples.

**Example 1.3.** Assumption V2 is fulfilled in the following cases:

i) The \(Y_n\) have arbitrary Poisson-distributions;

ii) The \(Y_n\) have arbitrary linear fractional distributions;

iii) There is a constant \(c < \infty\) such that \(Y_n \leq c\) a.s. for all \(n \geq 1\).

For the proof of (iii), rewrite V2 as
\[
\mathbb{E}[Y_n(Y_n - 1)] \leq c \mathbb{E}[Y_n | Y_n > 0] \cdot \mathbb{E}[(Y_n - 1)^+] 
\]
arb and observe that \(\mathbb{E}[Y_n | Y_n > 0] \geq 1\).

Here comes the main result of this section.

**Theorem 1.1.** Let the branching process \(Z\) in a varying environment fulfill Assumption V1. Then the conditions

i) \(q = 1\);

ii) \(\mathbb{E}[Z_n]^2 = o(\mathbb{E}[Z_n^2])\) as \(n \to \infty\);
\[\sum_{k=1}^{\infty} \frac{\rho_k}{f_1 \cdots f_{k-1}} = \infty.\]

are equivalent.

Condition (ii) can be equivalently expressed as \[E[Z_n] = o(\sqrt{\text{Var}[Z_n]}).\] Thus, shortly speaking, under Assumption \(V2\) we have a.s. extinction whenever the random fluctuations dominate the mean behavior of the process in the long run.

For the proof, we introduce a method of handling the measures \(f_{0,n}\), which will be useful elsewhere, too. It mimics the calculation we got to know for linear fractional distributions. For a probability measure \(f \in \mathcal{P}(\mathbb{N}_0)\) with mean \(0 < \bar{f} < \infty\), we define the function

\[\varphi_f(s) := \frac{1}{1-f(s)} = \frac{1}{f \cdot (1-s)}, \quad 0 \leq s < 1.\]

We also set

\[\varphi_f(1) := \lim_{s \to 1} \varphi_f(s) = \frac{f''(1)}{2f'(1)^2} = \frac{\tilde{f}}{2},\]

where the limit arises by means of the Taylor expansion

\[f(s) = 1 + f'(1)(s-1) + \frac{1}{2} f''(t)(s-1)^2 \quad \text{with some } t \in (s, 1).\]

From the convexity of the function \(f(s)\) we get that \(\varphi_f(s) \geq 0\) for all \(0 \leq s \leq 1\).

Then for probability measures \(f_1, \ldots, f_n\) with positive, finite means, we obtain

\[\frac{1}{1-f_{0,n}(s)} = \frac{1}{f_1 \cdot (1-f_{1,n}(s))} + \varphi_{f_1}(f_{1,n}(s)).\]

Iterating the formula and having in mind the conventions \(f_{n,n}(s) = s\) and \(\bar{f}_1 \cdots \bar{f}_{k-1} = 1\) for \(k = 1\), we arrive at the following expansion.

**Proposition 1.3.** For probability measures \(f_1, \ldots, f_n\) with positive, finite means \(\bar{f}_1, \ldots, \bar{f}_n\) we have

\[\frac{1}{1-f_{0,n}(s)} = \frac{1}{f_1 \cdots f_n \cdot (1-s)} + \sum_{k=1}^{n} \frac{\varphi_{f_k}(f_{k,n}(s))}{f_1 \cdots f_{k-1}}, \quad 0 \leq s < 1.\]

As seen from [1.2], the functions \(\varphi_{f_k}\) are constant for linear fractional probability measures. In general, we have the following sharp estimates.
Proposition 1.4.– Let $f \in \mathcal{P}(\mathbb{N}_0)$ with mean $0 < \bar{f} < \infty$. Then, it follows for $0 \leq s \leq 1$

$$
\frac{1}{2} \phi_f(0) \leq \phi_f(s) \leq 2 \phi_f(1).
$$

[1.10]

Note that $\phi_f$ is identical to zero if $f[z] = 0$ for all $z \geq 2$. Otherwise $\phi_f(0) > 0$, and the lower bound of $\phi_f$ becomes strictly positive. Choosing $s = 1$ and $s = 0$ in [1.10], we obtain $\phi_f(0)/2 \leq \phi_f(1)$ and $\phi_f(0) \leq 2 \phi_f(1)$. Note that for $f = \delta_k$ (Dirac-measure at point $k$) and $k \geq 2$, we have $\phi_f(1) = \phi_f(0)/2$, implying that the constants 1/2 and 2 in [1.10] cannot be improved.

Proof.– i) We prepare the proof by showing for $g_1, g_2 \in \mathcal{P}(\mathbb{N}_0)$ the following statement: If $g_1$ and $g_2$ have the same support and if, for any $k \geq 0$ with $g_1[k] > 0$, we have

$$
\frac{g_1[z]}{g_1[k]} \leq \frac{g_2[z]}{g_2[k]}
$$

for all $z > k$, then $g_1 \leq g_2$. Indeed, for $g_1[k] > 0$

$$
\frac{\sum_{z \geq k} g_1[z]}{1 - \sum_{z \geq k} g_1[z]} = \frac{\sum_{z \geq k} g_1[z]/g_1[k]}{\sum_{z < k} g_1[z]/g_1[k]} \leq \frac{\sum_{z \geq k} g_2[z]/g_2[k]}{\sum_{z < k} g_2[z]/g_2[k]} = \frac{\sum_{z \geq k} g_2[z]}{1 - \sum_{z \geq k} g_2[z]}
$$

and consequently

$$
\sum_{z \geq k} g_1[z] \leq \sum_{z \geq k} g_2[z].
$$

It follows that this inequality holds for all $k \geq 0$, since vanishing summands on the left-hand side may be removed. Summing the inequality over $k \geq 0$, we arrive at the claim.

For a special case, consider for $0 \leq s \leq 1$ and $r \in \mathbb{N}_0$ the probability measures

$$
g_s[z] := \frac{s^{r-z}}{1 + s + \cdots + s^r}, \quad 0 \leq z \leq r.
$$

Then for $0 < s \leq t$, $k \leq r$, $z > k$, we have $g_s[z]/g_s[k] \geq g_t[z]/g_t[k]$. We therefore obtain that

$$
g_s = \frac{s^{r-1} + 2s^{r-2} + \cdots + r}{1 + s + \cdots + s^r}$$
is a decreasing function in $s$. Moreover, $\bar{g}_0 = r$ and $\bar{g}_1 = r/2$, and it follows for $0 \leq s \leq 1$

\[
\frac{r}{2} \leq \frac{r + (r - 1)s + \cdots + sr^{r-1}}{1 + s + \cdots + s^r} \leq r.
\]  

[1.11]

ii) Next, we derive a second representation for $\varphi = \varphi_f$. We have

\[
1 - f(s) = \sum_{z=1}^{\infty} f[z](1 - s^z) = (1 - s) \sum_{z=1}^{\infty} f[z] \sum_{k=0}^{z-1} s^k,
\]

and

\[
f'(1)(1 - s) - (1 - f(s)) = (1 - s) \sum_{z=1}^{\infty} f[z] \sum_{k=0}^{z-1} (1 - s^k)
\]

\[
= (1 - s)^2 \sum_{z=1}^{\infty} f[z] \sum_{k=1}^{z-1} \sum_{j=0}^{k-1} s^j
\]

\[
= (1 - s)^2 \sum_{z=1}^{\infty} f[z][(z - 1) + (z - 2)s + \cdots + s^{z-2}].
\]

Therefore,

\[
\varphi(s) = \frac{f'(1)(1 - s) - (1 - f(s))}{f'(1)(1 - s) - (1 - f(s))}
\]

\[
= \frac{\sum_{z=1}^{\infty} f[z][(z - 1) + (z - 2)s + \cdots + s^{z-2}]}{f \cdot \sum_{k=1}^{\infty} f[k](1 + s + \cdots + s^{k-1})}.
\]

From [1.11], it follows

\[
\varphi(s) \leq \frac{\psi(s)}{f} \leq 2\varphi(s)
\]

[1.12] with

\[
\psi(s) := \frac{\sum_{z=1}^{\infty} f[z](z - 1)(1 + s + \cdots + s^{z-1})}{\sum_{k=1}^{\infty} f[k](1 + s + \cdots + s^{k-1})}.
\]

Now consider the probability measures $g_s \in \mathcal{P}(\mathbb{N}_0)$, $0 \leq s \leq 1$, given by

\[
g_s[z] := \frac{f[z+1](1 + s + \cdots + s^z)}{\sum_{k=0}^{\infty} f[k+1](1 + s + \cdots + s^k)}, \quad z \geq 0.
\]
Then for \( f[k + 1] > 0 \) and \( z > k \), after some algebra,

\[
gs[z] = \frac{f[z + 1]}{f[k + 1]} \prod_{v=1}^{z-k} \left( 1 + \frac{1}{s^{-1} + \cdots + s^{-k-v}} \right),
\]

which is an increasing function in \( s \). Therefore,

\[
\psi(s) = \bar{g}_s
\]
is increasing in \( s \). In combination with [1.12], we get

\[
\varphi(s) \leq \frac{\psi(s)}{f} \leq \frac{\psi(1)}{f} \leq 2 \varphi(1), \quad 2 \varphi(s) \geq \frac{\psi(s)}{f} \geq \frac{\psi(0)}{f} \geq \varphi(0).
\]

This gives the claim of the proposition. \( \square \)

**Proof (Proof of Theorem 1.1).**– If \( q = 1 \), then \( \Pr(\theta > n) \to 0 \) as \( n \to \infty \). Therefore, the implications (i) \( \Rightarrow \) (ii) and (ii) \( \Rightarrow \) (iii) follow from formula [1.8]. For the remaining part of the proof, note that \( \Pr(\theta > n) = \Pr(Z_n \neq 0) = 1 - f_{0,n}(0) \), such that it follows from Proposition 1.3

\[
\frac{1}{\Pr(\theta > n)} = \frac{1}{\bar{f}_1 \cdots \bar{f}_n} + \sum_{k=1}^{n} \frac{\varphi_{f_k}(f_{k,n}(0))}{\bar{f}_1 \cdots \bar{f}_{k-1}}.
\]

Moreover, we observe that \( V^2 \) reads

\[
\tilde{f}_n(\bar{f}_n)^2 \leq c \tilde{f}_n \cdot \frac{\bar{f}_n - (1 - f_n[0])}{1 - f_n[0]},
\]

which can be converted to \( \varphi_{f_n}(1) \leq c \varphi_{f_n}(0) \). Then Proposition 1.4 together with \( \varphi_{f_n}(1) = \tilde{f}_n/2 \) yields for \( 0 \leq s \leq 1 \)

\[
\frac{\tilde{f}_n}{4c} \leq \varphi_{f_n}(s).
\]

Together with [1.13] and Proposition 1.2, this implies with \( d := \max(1, 4c) \)

\[
\frac{1}{\Pr(\theta > n)} \geq \frac{1}{\bar{f}_1 \cdots \bar{f}_n} + \frac{1}{4c} \sum_{k=1}^{n} \frac{\tilde{f}_k}{\bar{f}_1 \cdots \bar{f}_{k-1}}
\]

\[
\geq \frac{1}{d} \frac{\mathbb{E}[Z_n(Z_n - 1)] + \mathbb{E}[Z_n]}{\mathbb{E}[Z_n]^2} = \frac{1}{d} \left( \frac{\text{Var}[Z_n]}{\mathbb{E}[Z_n]^2} + 1 \right).
\]

Now the implication (iii) \( \Rightarrow \) (i) follows from equation [1.6]. \( \square \)
1.3. Almost sure convergence

There are a few supermartingales which allow convergence considerations for branching processes $Z$ in a varying environment. Under Assumption V1, an obvious choice is the process $W = \{W_n, n \geq 0\}$, given by

$$W_n := \frac{Z_n}{f_1 \cdots f_n}, n \in \mathbb{N}_0,$$

which is easily seen to be a non-negative martingale. Therefore, there is an integrable random variable $W \geq 0$, such that

$$Z_n \xrightarrow{\text{a.s.}} W$$

as $n \to \infty$.

**Theorem 1.3.** For a branching process $Z$ with $Z_0 = 1$ and in a varying environment fulfilling the assumptions V1 and V2, we have

i) If $q = 1$ then $W = 0$ a.s.

ii) If $q < 1$ then $\mathbb{E}[W] = 1$.

**Proof.** The first claim is obvious. For the second one, we observe that $q < 1$ in view of Theorem 1.1 implies

$$\sum_{k=1}^{\infty} \frac{\rho_k}{f_1 \cdots f_{k-1}} < \infty.$$

From [1.6], it follows that

$$\sup_{n} \frac{\mathbb{E}[Z_n^2]}{\mathbb{E}[Z_n]^2} = \sup_{n} \frac{\text{Var}[Z_n]}{\mathbb{E}[Z_n]^2} + 1 < \infty.$$

Therefore, $W$ is a square-integrable martingale implying $\mathbb{E}[W] = \mathbb{E}[W_0] = 1$. □

The next theorem on the a.s. convergence of the unscaled process is remarkable, also in that it requires no assumptions at all. We name it the Church–Lindvall theorem. Among others, it clarifies as to which condition is needed for $Z$ with a positive probability to stick forever in some state $z \geq 1$. In its proof, we shall encounter a finer construction of a supermartingale.

**Theorem 1.4.** For a branching process $Z = \{Z_n, n \geq 0\}$ in a varying environment, there exists a random variable $Z_\infty$ with values in $\mathbb{N}_0 \cup \{\infty\}$ such that as $n \to \infty$

$$Z_n \xrightarrow{\text{a.s.}} Z_\infty.$$
Moreover,

\[ P(Z_\infty = 0 \text{ or } \infty) = 1 \iff \sum_{n=1}^{\infty} (1 - f_n[1]) = \infty. \]

**PROOF.**– i) We prepare the proof by showing that the sequence of probability measures \( f_{0,n} \) is vaguely converging to a (possibly defective) measure \( g \) on \( \mathbb{N}_0 \). Note that \( f_{0,n}[0] \to q \). Thus, either \( f_{0,n} \to q \delta_0 \) vaguely (with the Dirac measure \( \delta_0 \) at point 0), or else (by the Helly–Bray theorem) there exists a sequence of integers \( 0 = n_0 < n_1 < n_2 < \cdots \), such that, as \( k \to \infty \), we have \( f_{0,n_k} \to g \) vaguely with \( g \neq q \delta_0 \).

In the latter case, the limiting generating function \( g(s) \) is strictly increasing in \( s \), and \( f_{0,n_k}(s) \to g(s) \) for all \( 0 \leq s < 1 \). Then, given \( n \in \mathbb{N}_0 \), we define \( l_n := n_k, m_n := n_{k+1} \) with \( n_k \leq n < n_{k+1} \), thus \( l_n \leq n < m_n \). We want to show that \( f_{l_n,n} \) converges vaguely to \( \delta_1 \). For this purpose, we consider a subsequence \( n' \) such that both \( f_{l_n,n'} \) and \( f_{n',m_n} \) converge vaguely to measures \( h_1 \) and \( h_2 \). Going in \( f_{0,m_n} = f_{0,l_n} \circ f_{l_n,n'} \circ f_{n',m_n} \) to the limit, we obtain

\[ g(s) = g(h_1(h_2(s))), \quad 0 \leq s < 1. \]

Since \( g \) is strictly increasing, \( h_1(h_2(s)) = s \), which for generating functions implies \( h_1(s) = h_2(s) = s \). Thus, using the common sub-sub-sequence argument, \( f_{l_n,n} \to \delta_1 \) as \( n \to \infty \). It follows that, as \( n \to \infty \),

\[ f_{0,n}(s) = f_{0,l_n}(f_{l_n,n}(s)) \to g(s), \quad 0 \leq s < 1, \]

which means \( f_{0,n} \to g \) vaguely, as has been claimed.

ii) We now turn to the proof of the first statement. The case \( g(s) = 1 \) for all \( 0 \leq s < 1 \) is obvious, then \( g = \delta_0 \) and \( q = 1 \), and \( Z_n \) is a.s. convergent to 0. Thus, we are left with the case \( g(s) < 1 \) for all \( s < 1 \). Then, there is a decreasing sequence \( (b_n, n \geq 0) \) of real numbers, such that \( f_{0,n}(1/2) \leq b_n \leq 1 \) and \( b_n \downarrow g(1/2) \). We define the sequence \( (a_n, n \geq 0) \) using the following equation:

\[ f_{0,n}(a_n) = b_n. \]

Therefore, \( 1/2 \leq a_n \leq 1 \), and we also have \( f_{0,n+1}(a_{n+1}) \leq f_{0,n}(a_n) \) or equivalently \( f_{n+1}(a_{n+1}) \leq a_n \). Then, the process \( \mathcal{U} = \{U_n, n \geq 0\} \), given by

\[ U_n := a_n^{Z_n} \cdot I\{Z_n > 0\} \]
is a non-negative supermartingale. Indeed, because of $f_{n+1}(0)Z_n \geq I\{Z_n = 0\}$ and $f_{n+1}(a_{n+1}) \leq a_n$, we have
\[
E[U_{n+1} \mid Z_0, \ldots, Z_n] = f_{n+1}(a_{n+1})Z_n - f_{n+1}(0)Z_n \leq a_n Z_n - I\{Z_n = 0\} = U_n \text{ a.s.}
\]
Thus, $U_n$ is a.s. convergent to a random variable $U \geq 0$.

Now, we distinguish two cases. Either $g \neq q\delta_0$. Then $g(s)$ is strictly increasing, which implies $a_n \to 1/2$ as $n \to \infty$. Hence, the a.s. convergence of $U_n$ enforces the a.s. convergence of $Z_n$ with possible limit $\infty$.

Or $g = q\delta_0$. Then $g(1/2) = q$, implying that, for $n \to \infty$,
\[
E[U_n] = f_{0,n}(a_n) - f_{0,n}(0) = b_n - P(Z_n = 0) \to g(1/2) - q = 0
\]
and consequently $U = 0$ a.s. implying $U_n \to 0$ a.s. Since $a_n \geq 1/2$ for all $n$, this enforces that $Z_n$ converges a.s. to 0 or $\infty$. In both cases, $Z_n \to Z_\infty$ a.s. for some random variable $Z_\infty$.

iii) For the second statement, we use the representation $Z_n = \sum_{i=1}^{Z_{n-1}} Y_{i,n}$. Define the events $A_{z,n} := \{\sum_{i=1}^{Z_{n-1}} Y_{i,n} \neq z\}$. Then for $z \geq 1$
\[
P(A_{z,n}) \geq 3^{-z}(1 - f_n[1]).
\]
Indeed, if $f_n[1] \geq 1/3$, then
\[
P(A_{z,n}) \geq P(Y_{1,n} \neq 1, Y_{2,n} = \ldots = Y_{z,n} = 1)
\geq (1 - f_n[1])f_n[1]^z \geq 3^{-z}(1 - f_n[1]),
\]
and if $f_n[1] \leq 1/3$, then either $P(Y_{1,n} > 1) \geq 1/3$ or $P(Y_{1,n} = 0) \geq 1/3$ implying
\[
P(A_{z,n}) \geq P(\min(Y_{1,n}, \ldots, Y_{z,n}) > 1) + P(Y_{1,n} = \ldots = Y_{z,n} = 0)
\geq 3^{-z}(1 - f_n[1]).
\]
Now assume $\sum_{n=1}^{\infty}(1 - f_n[1]) = \infty$. As, for fixed $z$, the events $A_{z,n}$ are independent, it follows by the Borel–Cantelli lemma that these events occur a.s. infinitely often. From the a.s. convergence of $Z_n$, we get, for $z \geq 1$,
\[
P(Z_\infty = z) = P(Z_n \neq z \text{ finitely often}) \leq P(A_{z,n} \text{ occurs finitely often}) = 0.
\]
This implies that $P(1 \leq Z_\infty < \infty) = 0$. 


Conversely, let \( \sum_{n=1}^{\infty} (1 - f_n[1]) < \infty \). Then, for \( z \geq 1 \), with \( P(Z_0 = z) > 0 \), we have
\[
P(Z_\infty = z) \geq P(Z_n = z \text{ for all } n) \geq P(Z_0 = z) \left( \prod_{n=1}^{\infty} f_n[1] \right)^z > 0,
\]
and it follows that \( P(1 \leq Z_\infty < \infty) > 0 \). Hence, the proof is finished. \( \square \)

1.4. Family trees

Now we turn to family relations within populations. For this purpose, we introduce a labeling of individuals, the so-called \textit{Ulam–Harris labeling}, from which ancestral relationships become directly apparent. In this approach, individuals are identified by elements \( i \) of
\[
\mathcal{I} := \bigcup_{n=0}^{\infty} \mathbb{N}^n,
\]
which are written as finite strings \( j_1 \ldots j_n \) of positive integers. We agree that the set \( \mathbb{N}^0 \) consists of just the empty string \( \emptyset \). For two strings \( i, i' \), we write \( ii' \) for the concatenated string.

The labeling takes into account the generation \( g(i) \) and the number \( y(i) \) of offspring of each individual \( i \in \mathcal{I} \). If \( g(i) = n \), then \( i \) equals a string \( j_1 \ldots j_n \) of length \( n \). If \( i \) has \( y = y(i) \geq 1 \) children, then these are labeled as \( ij, 1 \leq j \leq y \). Accordingly, in case \( n \geq 1 \), the predecessor of \( i = j_1 \ldots j_n \) equals the truncated string \( i' = j_1 \ldots j_{n-1} \). We point out that the offspring of any individual gets ordered in this approach (whether with respect to age or otherwise remains unregarded). Certainly, the offspring numbers \( y(i) \) may also take the value zero.

An entire population is now captured by a subset \( t \) of \( \mathcal{I} \). We require the following properties:

i) \( \emptyset \in t \);

ii) for \( i \in \mathcal{I}, j \in \mathbb{N} \) we have \( ij \in t \Rightarrow i \in t \);

iii) for \( i \in \mathcal{I}, j \in \mathbb{N} \) we have \( ij \in t \Rightarrow ij' \in t \) for all \( 1 \leq j' \leq j \);

iv) for \( i \in t \) there is a \( j \in \mathbb{N} \) such that \( ij \notin t \).

The meaning of these items is easily understood. In view of (i), the empty string \( \emptyset \) represents the founding ancestor, which is the only individual in generation 0; thus, we implicitly assume that the size of generation zero is equal to 1. This founding ancestor is, according to (ii), the only individual lacking a predecessor within \( t \). Item (iii) takes
up the above labeling rule, and (iv) says that the sizes $y(i)$ of the offspring of $i$ are always finite. In terms of graph theory, we are dealing with rooted, ordered, locally finite trees. We call them family trees. When visualizing them in the plane, we place the root $\emptyset$ to the bottom and order siblings from left to right, as done in Figure 1.1. Therefore, trees are embedded into the plane in an essentially unique way.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.1.png}
\caption{A rooted tree}
\end{figure}

If a set $t \subset I$ satisfies the above properties (i) to (iv), we name it briefly a tree. Its height $h(t)$ is the maximal generation of its individuals,

$$h(t) := \max\{g(i) : i \in t\},$$

it can be finite or infinite. The generation sizes $z_n(t)$ of $t$ are

$$z_n(t) := \#\{i \in t : g(i) = n\}, \quad n \in \mathbb{N}_0,$$

they are finite for locally finite trees. We write $t \overset{h}{=} t'$ for trees $t, t'$ and integers $h \geq 0$, if $t$ and $t'$ coincide up to height $h$, that is, if

$$t \cap \bigcup_{n=0}^{h} \mathbb{N}^n = t' \cap \bigcup_{n=0}^{h} \mathbb{N}^n.$$

Next, in order to introduce random trees, we endow the set $\mathcal{T}$ of all trees $t \subset I$ with a $\sigma$-field. The natural choice is to consider the $\sigma$-field generated by all sets of the form $\{t' \in \mathcal{T} : t' \overset{h}{=} t\}$ with an integer $h \geq 0$ and a tree $t$. It is a straightforward exercise to show that all functionals of trees considered in this section, like $t \mapsto h(t)$ or $t \mapsto z_n(t)$, become measurable mappings this way. We are now ready to define a
branching tree in a varying environment $v = (f_1, f_2, \ldots)$. It is a $T$-valued random variable $T$ with a distribution characterized by

$$
P(T \overset{h}{=} t) = \prod_{i \in t : g(i) < h} f_{g(i)+1}[y(i)]$$

for any integer $h \geq 0$ and any tree $t$ (with generation and offspring numbers $g(i)$ and $y(i)$). It is a consequence of the Kolmogorov consistency theorem and the uniqueness theorem for probability measures that the distribution of $T$ is therefore well defined.

If, for $n \in \mathbb{N}_0$, we let

$$Z_n := z_n(T),$$

we are back to branching processes in a varying environment.

With this construction in hand, we may deal with new aspects of branching trees. We will not elaborate immediately but focus for later use on two tree constructions building some tree $t$ along a distinguished path (or spine). Such a path of length $\ell$ runs away from the root through the tree. It is given by a sequence $d_1, d_2, \ldots$ of natural numbers possessing the property that the strings $d_1 \ldots d_k$ belong to $t$ for all $k \leq \ell$. Its length $\ell$ obviously fulfills the inequality $\ell \leq h(t)$ and it may be finite or infinite.

For the first construction, we fix a natural number $n$. We shall give an enlightening description of a branching tree $T$ in varying environment $v = (f_1, f_2, \ldots)$ conditioned on the event that $Z_n > 0$. It can be observed that we may divide any tree $t$ into the subtrees $t_1, \ldots, t_y$ stemming from the individuals $1, \ldots, y$ of generation one (thus $y = y(\emptyset)$). Formally, these subtrees can be written as

$$t_j := \{i \in I : j_i \in t\}, \quad 1 \leq j \leq y.$$

Equally we may split the branching tree $T$ into subtrees $T_1, \ldots, T_y$. Then with $h' := h - 1$, $h \geq 1$,

$$\{T \overset{h}{=} t\} = \{Y = y, T_1 \overset{h'}{=} t_1, \ldots, T_y \overset{h'}{=} t_y\}$$

and, due to independence,

$$P_v(T \overset{h}{=} t) = f_1[y] \prod_{j=1}^{y} P_{v_1}(T_j \overset{h'}{=} t_j).$$

Note that the right-hand probabilities refer to the shifted environment $v_1 := (f_2, f_3, \ldots)$. 
Next, if \( t \) has a height of at least \( n \), then there is a distinguished individual \( d \) in generation one, \( 1 \leq d \leq z \), such that \( t_d \) has height at least \( n - 1 \) but the trees \( t_j \), \( 1 \leq j < d \) have heights less than \( n - 1 \). Accordingly, we dissect the previous formula for \( h \geq n \) as

\[
P_v(T \overset{h}{=} t) = f_1[y] \prod_{j=1}^{d-1} P_{v_1}(T_j = t_j) \times P_{v_1}(T_d \overset{h'}{=} t_d) \times \prod_{j=d+1}^y P_{v_1}(T_j \overset{h'}{=} t_j).
\]

In order to pass to conditional probabilities, we reweight the right-hand terms. Letting for \( 1 \leq d \leq y < \infty \)

\[
g_n[d, y] := \frac{1 - f_1,n[0]}{1 - f_0,n[0]} f_1[y] f_1,n[0]^{d-1}
\]

and recalling that \( P_v(Z_n = 0) = f_0,n[0] \) and \( P_{v_1}(Z_{n-1} = 0) = f_1,n[0] \), we arrive at the equation

\[
P_v(T \overset{h}{=} t \mid Z_n > 0) = g_n[d, y] \times \prod_{j=1}^{d-1} P_{v_1}(T_j = t_j \mid Z_{n-1} = 0)
\]

\[
\times P_{v_1}(T_d \overset{h'}{=} t_d \mid Z_{n-1} > 0) \times \prod_{j=d+1}^y P_{v_1}(T_j \overset{h'}{=} t_j).
\]

Moreover,

\[
\sum_{1 \leq d \leq y < \infty} g_n[d, y] = \frac{1}{1 - f_0,n[0]} \sum_{y \geq 1} f_1[y] (1 - f_1,n[0]^y) = \frac{1 - f_1(f_1,n[0])}{1 - f_0,n[0]} = 1.
\]

Thus, we may regard \( g_n \) as a probability measure and reinterpret the above formulas as in the following lemma.

**Lemma 1.1.**—Let \( n \in \mathbb{N} \) and let \( v = (f_1, f_2, \ldots) \) be a varying environment. Let the random tree \( T \) be composed of the subtrees \( T_1, \ldots, T_Y \) in generation one and let \( D \) be an integer-valued random variable with \( 1 \leq D \leq Y \). Assume that:

i) the distribution of \((D, Y)\) has the weights \( g_n[d, y], 1 \leq d \leq y < \infty\);

ii) given \((D, Y)\) the random trees \( T_1, \ldots, T_Y \) are independent branching trees in environment \( v_1 = (f_2, f_3, \ldots) \), in which \( T_1, \ldots, T_{D-1} \) are conditioned to have heights less than \( n - 1 \), \( T_D \) is conditioned to have a height of at least \( n - 1 \) and \( T_{D+1}, \ldots, T_Y \) are unconditioned.

Then \( T \) is a branching random tree in environment \( v \) conditioned to have a height of at least \( n \).
A crucial point of this lemma is that it can be reapplied to the tree $T_D$ growing at the distinguished individual $D$, now with $n$ replaced by $n - 1$. This procedure may be iterated several times in which the conditioned tree $T$ arises in an $n$-stage construction.

As a byproduct a distinguished path within $T$ arises given by the numbers $D_1, D_2, \ldots, D_n$ and the distinguished individuals $\Delta_1, \ldots, \Delta_n$. It ends in the ‘leftmost’ individual $\Delta_n$ in generation $n$, that is, in the string $\Delta_n = D_1 \ldots D_n$ appearing leftmost in the visualization described above. (Another way to phrase this is to say that the string $D_1 \ldots D_n$ is smallest in the lexicographical order among all strings $j_1 \ldots j_n$ of length $n$ within $T$.)

This suggests the need to upturn our construction, to start from the distinguished path and then to attach step by step trees to its left and right. Consider the probability measures $g_{mn}, 1 \leq m \leq n$, with weights

$$
g_{mn}[d, y] := \frac{1 - f_{m,n}[0]}{1 - f_{m-1,n}[0]} f_m[y] f_{m,n}[0]^{d-1}, \quad 1 \leq d \leq y < \infty. \quad [1.14]
$$

Then the following procedure results in a branching tree $T$ in varying environment $v = (f_1, f_2, \ldots)$ conditioned on the event $\{Z_n > 0\}$. We call it a Geiger tree.

![Figure 1.2. A Geiger tree](image-url)
1.4.1. Construction of the Geiger tree

1) Establish the distinguished path containing the founding ancestor $\emptyset$ and $n$ distinguished individuals in generations 1 to $n$.

2) Generate independent random variables $(D_1, Y_1), \ldots, (D_n, Y_n)$ with distributions $g_{1n}, \ldots, g_{nn}$. Supply the distinguished individual in generation $m$ with $Y_m - 1$ siblings, of these $D_m - 1$ ‘younger’ to the left and $Y_m - D_m$ ‘older’ to the right.

3) Given these random variables, we generate independent branching trees $T_{1,m}, \ldots, T_{D_m-1,m}$ and $T_{D_m+1,m}, \ldots, T_{Y_m,m}$ in the varying environment $v_m := (f_{m+1}, f_{m+2}, \ldots)$, $1 \leq m \leq n$, such that the first are conditioned to have heights less than $n - m$ and the second are unconditioned. Attach these trees to the siblings in generation $m$, the conditioned trees to the left and the unconditioned trees to the right of the distinguished path.

4) Complete the tree by adding an independent, unconditioned branching tree $T_{D_n,n}$ in environment $v_n = (f_{n+1}, f_{n+2}, \ldots)$ on top of the distinguished individual in generation $n$.

Now, we turn to the other tree construction yielding a size-biased branching tree in varying environment $v = (f_1, f_2, \ldots)$. It arises from the Geiger tree by letting $n$ go to $\infty$, provided that the underlying branching process gets a.s. extinct. Then, as $n \to \infty$, $P_{v_m}(Z_n = 0) \to 1$ and

$$P_{v_m}(T \overset{h}{=} t \mid Z_n = 0) \to P_{v_m}(T \overset{h}{=} t)$$

for any $m \geq 0$, that is, the conditioned trees get unconditioned in the limit. Moreover, we have $f_{m,n}[0] \to 1$ and for $m \geq 1$

$$\frac{1 - f_{m,n}[0]}{1 - f_{m-1,n}[0]} = \frac{1 - f_{m,n}[0]}{f_m(1) - f_m(f_{m,n}[0])} \to \frac{1}{f'_m(1)}$$

implying

$$g_{mn}[d, y] \to g_{m\infty}[d, y] := \frac{1}{f_m} f_m[y].$$

Again

$$\sum_{1 \leq d \leq y < \infty} g_{m\infty}[d, y] = \sum_{y \geq 1} \frac{1}{f_m} y f_m[y] = 1.$$
Now the pair \((D_m, Y_m)\) can be obtained in a particularly lucid manner: first, generate the random variable \(Y_m\) with the \textit{size-biased distribution} \(f_m^*\) having weights
\[
f_m^*[y] := \frac{1}{f_m} y f_m[y], \quad y \geq 1. \tag{1.15}
\]
Second, choose \(D_m\) uniformly among the values of \(1, \ldots, Y_m\). Then, indeed,
\[
P(Y_m = y, D_m = d) = f_m^*[y] \cdot \frac{1}{y} = g_m \infty[d, y].
\]
Note that the bias to the left of the Geiger tree vanishes in the limit \(n \to \infty\). We denote the limiting tree by \(T^*\).

### 1.4.2. Construction of the size-biased tree \(T^*\)

1) Establish the distinguished path containing the founding ancestor \(\emptyset\) and a path of distinguished individuals of infinite length.

2) Generate independent random variables \(Y_1, Y_2, \ldots\) with distributions \(f_1^*, f_2^*, \ldots\). Given their values \(y_1, y_2, \ldots\) let \(D_1, D_2, \ldots\) be independent variables such that \(D_m\) is uniformly distributed in \(\{1, \ldots, y_m\}\). Supply the distinguished individual in generation \(m\) with \(Y_m - 1\) siblings, of these \(D_m - 1\) to the left and \(Y_m - D_m\) to the right.

3) Given these random variables generate independent (unconditioned) branching trees \(T_{1,m}, \ldots, T_{D_m-1,m}\) and \(T_{D_m+1,m}, \ldots, T_{Y_m,m}\) in environment \(v_m = (f_{m+1}, f_{m+2}, \ldots)\), \(m \geq 1\), and attach these trees to the siblings in generation \(m\) to the left and right of the distinguished path.

We point out that this construction is meaningful not only in the case where the corresponding branching process \(Z\) gets a.s. extinct, but also in the opposite situation. It is this case where we shall use the tree \(T^*\) later. The name \textit{size-biased tree} comes from the following crucial property. Recall that the distinguished path in \(T^*\) is made up of the strings \(\Delta_n := D_1 \ldots D_n, n \geq 1\).

**Lemma 1.2.**—Let \(t\) be a tree of a height of at least \(n \geq 1\) and let \(d \in t\) be an individual in generation \(n\). Then for a branching tree \(T\) and a size-biased branching tree \(T^*\), both in the same environment \(v = (f_1, f_2, \ldots)\),
\[
P(T^* \overset{n}{=} t, \Delta_n = d) = \frac{1}{f_1 \cdots f_n} P(T \overset{n}{=} t). \tag{1.16}
\]
Equivalently,
\[
P(T^* \overset{n}{=} t) = \frac{z_n(t)}{f_1 \cdots f_n} P(T \overset{n}{=} t),
\]
and given \(Z_n\), the distinguished individual \(\Delta_n\) in \(T^*\) is uniformly distributed among all individuals in generation \(n\).
PROOF. – We proceed by induction. For \( n = 1 \), we have

\[
P(T^* = t, \Delta_1 = d) = \frac{1}{f_1} f_1[y] = \frac{1}{f_1} P(T = t),
\]

where \( y = y(\emptyset) \) is the number of offspring of \( \emptyset \) within \( t \). For the inductive step, let the tree \( t \) have a height of at least \( n + 1 \) and let \( d \in t \) with \( g(d) = n + 1 \). Denote by \( d' \) the predecessor of \( d \). Then, by construction of the size-biased tree and by the induction hypothesis, we have

\[
P(T^* n+1 = t, \Delta_{n+1} = d) = P(T^* n = t, \Delta_n = d') \frac{1}{f_{n+1}} f_{n+1}[y(d')] \prod_{i \in t : i \neq d', g(i) = n} f_{n+1}[y(i)]
\]

\[
= \frac{1}{f_1 \cdots f_n} P(T^* n = t) \frac{1}{f_{n+1}} \prod_{i \in t : g(i) = n} f_{n+1}[y(i)]
\]

\[
= \frac{1}{f_1 \cdots f_{n+1}} P(T^* n+1 = t).
\]

The proof is complete. \( \square \)

1.5. Notes

Branching processes in a varying environment have been studied since the early 1970s starting with papers of Church [CHU 71], Fearn [FEA 71] and Athreya and Karlin [ATH 71a]. The question of a.s. extinction was first addressed by Agresti [AGR 75] and then further investigated by Jirina [JIR 76], Fujimari [FUJ 80], Lyons [LYO 92, HU 11] and others. The above Theorem 1.1 is taken from Kersting [KER 17].

Soon it became clear that, in comparison with Galton–Watson processes, new phenomena appear. Thus, building on the work of Church, Lindvall [LIN 74] obtained the above a.s. convergence result Theorem 1.4, clarifying that a branching process in a varying environment may stick to positive states forever with a positive probability. Later, MacPhee and Schuh [MAC 83] discovered that a branching process in a varying environment may diverge at different exponential rates (see also [D’SO 94, D’SO 92, GOE 76]).
The paper of Jagers [JAG 74] is the first to deal with the problem of the classification of branching processes in a varying environment. Later, this question was marginally studied and only recently answered in sufficient generality in [KER 17] (see also [BHA 17]).

The estimate from above as given in Proposition 1.4 was first established (with a different proof) by Agresti [AGR 75], Lemma 2, see also [KOZ 76, GEI 00]. The estimate from below is taken from [KER 17]. Precise estimates for $\varphi_f$ from above and below, covering also the multidimensional case and based on moments higher than 2, are obtained in a very interesting paper by Zubkov [ZU 94].

Some estimates from above and below for $P(\theta < \infty)$ and $E[\theta]$ are contained in [YU 09].