MATHEMATICS AND STATISTICS SERIES

BRANCHING PROCESSES, BRANCHING RANDOM WALKS AND BRANCHING PARTICLE FIELDS SET

Volume 1

Discrete Time Branching Processes in Random Environment

Götz Kersting
Vladimir Vatutin

ISTE

WILEY

Discrete Time Branching Processes in Random Environment

Branching Processes, Branching Random Walks and Branching Particle Fields Set

coordinated by
Elena Yarovaya

Volume 1

Discrete Time Branching Processes in Random Environment

Götz Kersting
Vladimir Vatutin

iSTE

WILEY

First published 2017 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Control Number: 2017949423

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-78630-252-6

Contents

Preface ix

List of Notations xiii

Chapter 1. Branching Processes in Varying Environment 1

1.1. Introduction 1
1.2. Extinction probabilities 6
1.3. Almost sure convergence 12
1.4. Family trees 15
1.4.1. Construction of the Geiger tree 20
1.4.2. Construction of the size-biased tree T^* 21
1.5. Notes 22

Chapter 2. Branching Processes in Random Environment 25

2.1. Introduction 25
2.2. Extinction probabilities 28
2.3. Exponential growth in the supercritical case 30
2.4. Three subcritical regimes 35
2.5. The strictly critical case 43
2.6. Notes 45

Chapter 3. Large Deviations for BPREs 47

3.1. Introduction 47
3.2. A tail estimate for branching processes in a
varying environment 50
3.3. Proof of Theorem 3.1 55
3.4. Notes 59

Chapter 4. Properties of Random Walks . . . 61

4.1. Introduction . . . 61
4.2. Sparre-Andersen identities . . . 62
4.3. Spitzer identity . . . 69
4.4. Applications of Sparre-Andersen and Spitzer identities . . . 70
4.4.1. Properties of ladder epochs and ladder heights . . . 70
4.4.2. Tail distributions of ladder epochs . . . 75
4.4.3. Some renewal functions . . . 76
4.4.4. Asymptotic properties of L_n and M_n . . . 79
4.4.5. Arcsine law . . . 83
4.4.6. Large deviations for random walks . . . 84
4.5. Notes . . . 90

Chapter 5. Critical BPREs: the Annealed Approach . . . 91

5.1. Introduction . . . 91
5.2. Changes of measures . . . 92
5.3. Properties of the prospective minima . . . 99
5.4. Survival probability . . . 107
5.5. Limit theorems for the critical case (annealed approach) . . . 116
5.6. Environment providing survival . . . 122
5.7. Convergence of $\log Z_n$ . . . 126
5.8. Notes . . . 127

Chapter 6. Critical BPREs: the Quenched Approach . . . 133

6.1. Introduction . . . 133
6.2. Changes of measures . . . 134
6.3. Probability of survival . . . 138
6.4. Yaglom limit theorems . . . 142
6.4.1. The population size at non-random moments . . . 143
6.4.2. The population size at moments $nt, 0 < t < 1$ . . . 149
6.4.3. The number of particles at moment $\tau(n) \leq nt$ . . . 150
6.4.4. The number of particles at moment $\tau(n) > nt$ . . . 157
6.5. Discrete limit distributions . . . 159
6.6. Notes . . . 167

Chapter 7. Weakly Subcritical BPREs . . . 169

7.1. Introduction . . . 169
7.2. The probability measures $\widehat{\mathbb{P}}^+$ and $\widehat{\mathbb{P}}^-$ . . . 172
7.3. Proof of theorems . . . 180
7.3.1. Proof of Theorem 7.1 . . . 188

7.3.2. Proof of Theorem 7.2 190
7.3.3. Proof of Theorem 7.3 191
7.4. Notes 195

Chapter 8. Intermediate Subcritical BPREs 197

8.1. Introduction 197
8.2. Proof of Theorems 8.1 to 8.3 201
8.3. Further limit results 209
8.4. Conditioned family trees 212
8.5. Proof of Theorem 8.4 219
8.6. Notes 225

Chapter 9. Strongly Subcritical BPREs 227

9.1. Introduction 227
9.2. Survival probability and Yaglom-type limit theorems 229
9.3. Environments providing survival and dynamics of the population size 234
9.3.1. Properties of the transition matrix P^* 240
9.3.2. Proof of Theorem 9.2 243
9.3.3. Proof of Theorem 9.3 248
9.4. Notes 251

Chapter 10. Multi-type BPREs 257

10.1. Introduction 257
10.2. Supercritical MBPREs 264
10.3. The survival probability of subcritical and critical MBPREs 264
10.4. Functional limit theorem in the critical case 266
10.5. Subcritical multi-type case 266

Appendix 269

Bibliography 275

Index 285

Preface

Branching processes constitute a fundamental part in the theory of stochastic processes. Very roughly speaking, the theory of branching processes deals with the issue of exponential growth or decay of random sequences or processes. Its central concept consists of a system or a population made up of particles or individuals which independently produce descendants. This is an extensive topic dating back to a publication of F. Galton and H. W. Watson in 1874 on the extinction of family names and afterwards dividing into many subareas. Correspondingly, it is treated in a number of monographs starting in 1963 with T. E. Harris' seminal *The Theory of Branching Processes* and supported in the middle of the 1970s by Sevastyanov's *Verzweigungsprozesse*, Athreya and Ney's *Branching Processes*, Jagers' *Branching Processes with Biological Applications* and others. The models considered in these books mainly concern branching processes evolving in a constant environment. Important tools in proving limit theorems for such processes are generating functions, renewal type equations and functional limit theorems.

However, these monographs rarely touch the matter of branching processes in a random environment (BPREs). These objects form not so much a subclass but rather an extension of the area of branching processes. In such models, two types of stochasticity are incorporated: on the one hand, demographic stochasticity resulting from the reproduction of individuals and, on the other hand, environmental stochasticity stemming from the changes in the conditions of reproduction along time. A central insight is that it is often the latter component that primarily determines the behavior of these processes. Thus, the theory of BPREs gains its own characteristic appearance and novel aspects appear such as a phase transition. From a technical point of view, the study of such processes requires an extension of the range of methods to be used in comparison with the methods common in the classical theory of branching processes. Other techniques should be attracted, in particular from the theory of random walks which plays an essential role in proving limit theorems for BPREs.

With this volume, we have two purposes in mind. First, we have to assert that the basics of the theory of BPREs are somewhat scattered in the literature (since the late 1960s), from which they are not at all easily accessible. Thus, we start by presenting them in a unified manner. In order to simplify matters, we confine ourselves from the beginning to the case where the environment varies in an i.i.d. fashion, the model going back to the now classical paper written by Smith and Wilkinson in 1969. We also put together that material which is required from the topic of branching processes in a varying environment. Overall, the proofs are now substantially simplified and streamlined but, at the same time, some of the theorems could be better shaped.

Second, we would like to advance some scientific work on branching processes in a random environment conditioned on survival which was conducted since around 2000 by a German-Russian group of scientists consisting of Valery Afanasyev, Christian Böinghoff, Elena Dyakonova, Jochen Geiger, Götz Kersting, Vladimir Vatutin and Vitali Wachtel. This research was generously supported by the German Research Association DFG and the Russian Foundation for Basic Research RFBR. In this book, we again do not aim to present our results in their most general setting, yet an ample amount of technical context cannot be avoided.

We start the book by describing in Chapter 1 some properties of branching processes in a varying environment (BPVEs). In particular, we give a (short) proof of the theorem describing the necessary and sufficient conditions for the dichotomy in the asymptotic behavior of a BPVE: such a process should either die or its population size should tend to infinity with time. Besides the construction of family trees, size-biased trees and Geiger's tree, representing conditioned family trees, are described here in detail. These trees play an important role in studying subcritical BPREs. Chapter 2 leads the reader in to the world of BPREs. It contains classification of BPREs, describes some properties of supercritical BPREs and gives rough estimates for the growth rate of the survival probability for subcritical BPREs. Conclusions of Chapter 2 are supported by Chapter 3 where the asymptotic behavior of the probabilities of large deviations for all types of BPREs is investigated.

Properties of BPREs are closely related to the properties of the so-called associated random walk (ARW) constituted by the logarithms of the expected population sizes of particles of different generations. This justifies the appearance of Chapter 4 that includes some basic results on the theory of random walks and a couple of findings concerning properties of random walks conditioned to stay non-negative or negative and probabilities of large deviations for different types of random walks.

Chapters 5 through 9 deal with various statements describing the asymptotic behavior of the survival probability and Yaglom-type functional conditional limit theorems for the critical and subcritical BPREs and analyzing properties of the ARW providing survival of a BPRE for a long time. Here, the theory of random walks conditioned to stay non-negative or negative demonstrates its beauty and power.

Thus, it is shown in Chapter 5 (under the annealed approach) that if a critical BPRE survives up to a distant moment n, then the minimum value of the ARW on the interval $[0, n]$ is attained at the beginning of the evolution of the BPRE and the longtime behavior of the population size of such BPREs (conditioned on survival) resembles the behavior of the ordinary supercritical Galton–Watson branching processes. If, however, a critical BPRE is considered under the quenched approach (Chapter 6) then, given the survival of the process for a long time, the evolution of the population size in the past has an oscillating character: the periods when the population size was very big were separated by intervals when the size of the population was small.

Chapters 7–9 are devoted to the weakly, intermediately and strongly subcritical BPREs investigated under the annealed approach. To study properties of such processes, it is necessary to make changes in the initial measures based on the properties of the ARWs. The basic conclusion of Chapters 7–8 is: the survival probability of the weakly and intermediately subcritical BPREs up to a distant moment n is proportional to the probability for the corresponding ARW to stay non-negative within the time interval $[0, n]$. Finally, it is shown in Chapter 9 that properties of strongly subcritical BPREs are, in many respect, similar to the properties of the subcritical Galton–Watson branching processes. In particular, the survival probability of such a process up to a distant moment n is proportional to the expected number of particles in the process at this moment.

We do not pretend that this book includes all interesting and important results established up to now for BPREs. In particular, we do not treat here BPREs with immigration and multitype BPREs. The last direction of the theory of BPREs is a very promising field of investigation that requires study of properties of Markov chains generated by products of random matrices. To attract the attention of future researchers to this field, we give a short survey of some recent results for multitype BPREs in Chapter 10. The book is concluded by an Appendix that contains statements of results used in the proofs of some theorems but not fitting the main line of the monograph.

Götz KERSTING
Vladimir VATUTIN
August 2017

List of Notations

$\mathcal{P}(\mathbb{N}_0)$	set of all probability measures f on $\mathbb{N}_0 = \{0,1,\ldots\}$, 2
$f[z]$, $f(s)$	weights and generating function of the measure f, 2
$\bar{f}$	the mean of the measure f, 2
$\tilde{f}$	the normalized second factorial moment, 2
f^*	size-biased measure, 21
$f_{m,n} := f_{m+1} \circ \cdots \circ f_n$	convolutions of probability measures, 3
$f_{m,n} := f_m \circ \cdots \circ f_{n+1}$	convolutions of probability measures, 107
$\varkappa(a,f) := (\bar{f}\,)^{-2} \sum_{y=a}^{\infty} y^2 f[y]$	truncated second moment, 102
θ	the moment of extinction of a branching process, 6, 28
T^*	size-biased tree, 21
$\mathcal{V}$	random environment, 25
$X := \log \overline{F}$	logarithm of expected population size, 27
$\kappa(\lambda) := \log \mathbb{E}[e^{\lambda X}]$	cumulant generating function, 36
γ	strict descending ladder epoch, 61
γ'	weak descending ladder epoch, 61
Γ	strict ascending ladder epoch, 61
Γ'	weak ascending ladder epoch, 62
$L_n := \min(S_0, S_1, \ldots, S_n)$	38
$M_n := \max(S_1, \ldots, S_n)$	76
$\tau(n) := \min\{0 \le k \le n : S_k = L_n\}$	the left-most moment when the minimum value of the random walk on the interval $[0,n]$ is attained, 70
$\mathbf{P}$	probability measure given the environment, 3, 26
$\mathbf{E}$	expectation given the environment, 3, 26

$\mathbb{P}$	probability measure obtained by averaging with respect to the environment, 25
$\mathbb{E}$	expectation taken after averaging with respect to the environment, 25
$\mathbb{P}^+, \mathbb{P}^-$	change of measure, 93
$\mathbb{P}^\pm$	change of measure, 134
$\widehat{\mathbb{P}}$	change of measure, 170, 198
$\widehat{\mathbb{P}}^+, \widehat{\mathbb{P}}^-$	change of measure, 172
$\mathbb{P}^*$	change of measure, 234

1

Branching Processes in Varying Environment

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.

Discrete Time Branching Processes in Random Environment, First Edition.
Götz Kersting and Vladimir Vatutin.

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*,

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

can be obtained from the generating functions as

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

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 $(\Omega, \mathcal{F}, \mathbf{P})$ 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 $\mathcal{Z} = \{Z_n, n \in \mathbb{N}_0\}$ with values in $\mathbb{N}_0$ is called a *branching process with environment* v, if for any integers $z \geq 0$, $n \geq 1$

$$\mathbf{P}(Z_n = z \mid Z_0, \ldots, Z_{n-1}) = (f_n^{Z_{n-1}})[z] \quad \mathbf{P}\text{-a.s.}$$

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

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

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

This corresponds to the following conception of the process $\mathcal{Z}$: 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. □

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

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

Moreover, let $f_{n,n}$ be the Dirac measure δ_1.

PROPOSITION 1.1.– *Let $\mathcal{Z}$ 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. □

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

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

We shall see that it is convenient to use the parameters $\bar{f}$ and $\widetilde{f}$ 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 δ_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

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

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

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

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

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

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

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

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

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

$$\frac{1}{1 - f_{0,n}(s)} = \frac{1}{\bar{f}_1 \cdots \bar{f}_n \cdot (1 - s)} + \frac{1}{2}\sum_{k=1}^{n} \frac{\widetilde{f}_k}{\bar{f}_1 \cdots \bar{f}_{k-1}}. \qquad [1.3]$$

It implies that the measure $f_{0,n}$ itself is of the linear fractional type with a mean and normalized second factorial moment

$$\bar{f}_{0,n} = \bar{f}_1 \cdots \bar{f}_n, \quad \widetilde{f}_{0,n} = \sum_{k=1}^{n} \frac{\widetilde{f}_k}{\bar{f}_1 \cdots \bar{f}_{k-1}}.$$

This property of perpetuation is specific for probability measures of the linear fractional type. $\square$

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

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

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 ∞).

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

$$\mathbf{E}[Z_n] = \bar{f}_1 \cdots \bar{f}_n, \ \frac{\mathbf{E}[Z_n(Z_n - 1)]}{\mathbf{E}[Z_n]^2} = \sum_{k=1}^{n} \frac{\widetilde{f}_k}{\bar{f}_1 \cdots \bar{f}_{k-1}}. \qquad [1.5]$$

We note that these equations entail the similarly built formula

$$\frac{\mathbf{Var}[Z_n]}{\mathbf{E}[Z_n]^2} = \sum_{k=1}^{n} \frac{\rho_k}{\bar{f}_1 \cdots \bar{f}_{k-1}}, \tag{1.6}$$

set up for the standardized variances

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

of the probability measures f_k. Indeed,

$$\begin{aligned}
\sum_{k=1}^{n} \frac{\rho_k}{\bar{f}_1 \cdots \bar{f}_{k-1}} &= \sum_{k=1}^{n} \frac{f_k''(1) + f_k'(1) - f_k'(1)^2}{\bar{f}_1 \cdots \bar{f}_{k-1} \cdot \bar{f}_k^2} \\
&= \sum_{k=1}^{n} \frac{f_k''(1)}{\bar{f}_1 \cdots \bar{f}_{k-1} \cdot \bar{f}_k^2} + \sum_{k=1}^{n} \Big(\frac{1}{\bar{f}_1 \cdots \bar{f}_k} - \frac{1}{\bar{f}_1 \cdots \bar{f}_{k-1}} \Big) \\
&= \sum_{k=1}^{n} \frac{\widetilde{f}_k}{\bar{f}_1 \cdots \bar{f}_{k-1}} + \frac{1}{\bar{f}_1 \cdots \bar{f}_n} - 1 \\
&= \frac{\mathbf{E}[Z_n(Z_n - 1)]}{\mathbf{E}[Z_n]^2} + \frac{1}{\mathbf{E}[Z_n]} - 1 = \frac{\mathbf{Var}[Z_n]}{\mathbf{E}[Z_n]^2}.
\end{aligned}$$

1.2. Extinction probabilities

For a branching process $\mathcal{Z}$, let

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

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

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

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

$$\mathbf{P}(\theta > n) = \mathbf{P}(Z_n \geq 1) \leq \mathbf{E}[Z_n]$$

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

$$\liminf_{n \to \infty} \bar{f}_1 \cdots \bar{f}_n = 0 \quad \Rightarrow \quad q = 1. \tag{1.7}$$

On the other hand, the Paley–Zygmund inequality tells us that

$$\mathbf{P}(\theta > n) = \mathbf{P}(Z_n > 0) \geq \frac{\mathbf{E}[Z_n]^2}{\mathbf{E}[Z_n^2]} = \frac{\mathbf{E}[Z_n]^2}{\mathbf{E}[Z_n] + \mathbf{E}[Z_n(Z_n - 1)]},$$

which in combination with equation [1.6] yields the bound

$$\frac{1}{\mathbf{P}(\theta > n)} \leq \frac{\mathbf{E}[Z_n^2]}{\mathbf{E}[Z_n]^2} = 1 + \frac{\mathbf{Var}[Z_n]}{\mathbf{E}[Z_n]^2} = 1 + \sum_{k=1}^{n} \frac{\rho_k}{\bar{f}_1 \cdots \bar{f}_{k-1}}. \qquad [1.8]$$

Thus, the question arises as to which one of both bounds captures the size of $\mathbf{P}(\theta > n)$ 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* $n \geq 1$

$$\mathbf{E}[Y_n(Y_n - 1)] \leq c\mathbf{E}[Y_n] \cdot \mathbf{E}[Y_n - 1 \mid Y_n > 0],$$

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.

For the proof of (iii), rewrite $V2$ as

$$\mathbf{E}[Y_n(Y_n - 1)] \leq c\mathbf{E}[Y_n \mid Y_n > 0] \cdot \mathbf{E}[(Y_n - 1)^+]$$

and observe that $\mathbf{E}[Y_n \mid Y_n > 0] \geq 1$. □

Here comes the main result of this section.

THEOREM 1.1.– *Let the branching process* $\mathcal{Z}$ *in a varying environment fulfill Assumption* $V1$*. Then the conditions*

i) $q = 1$;

ii) $\mathbf{E}[Z_n]^2 = o\big(\mathbf{E}[Z_n^2]\big)$ *as* $n \to \infty$;

iii) $\displaystyle\sum_{k=1}^{\infty} \frac{\rho_k}{\bar{f}_1 \cdots \bar{f}_{k-1}} = \infty.$

are equivalent.

Condition (ii) can be equivalently expressed as $\mathbf{E}[Z_n] = o\big(\sqrt{\mathbf{Var}[Z_n]}\big)$. 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

$$\varphi_f(s) := \frac{1}{1 - f(s)} - \frac{1}{\bar{f} \cdot (1 - s)}, \quad 0 \le s < 1.$$

We also set

$$\varphi_f(1) := \lim_{s \to 1} \varphi_f(s) = \frac{f''(1)}{2f'(1)^2} = \frac{\widetilde{f}}{2}, \qquad [1.9]$$

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 $\varphi_f(s) \ge 0$ for all $0 \le s \le 1$.

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

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

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*

$$\frac{1}{1 - f_{0,n}(s)} = \frac{1}{\bar{f}_1 \cdots \bar{f}_n \cdot (1 - s)} + \sum_{k=1}^{n} \frac{\varphi_{f_k}(f_{k,n}(s))}{\bar{f}_1 \cdots \bar{f}_{k-1}}, \quad 0 \le s < 1.$$

As seen from [1.2], the functions φ_{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* $0 \le s \le 1$

$$\frac{1}{2}\varphi_f(0) \le \varphi_f(s) \le 2\varphi_f(1). \qquad [1.10]$$

Note that φ_f is identical to zero if $f[z] = 0$ for all $z \ge 2$. Otherwise $\varphi_f(0) > 0$, and the lower bound of φ_f becomes strictly positive. Choosing $s = 1$ and $s = 0$ in [1.10], we obtain $\varphi_f(0)/2 \le \varphi_f(1)$ and $\varphi_f(0) \le 2\varphi_f(1)$. Note that for $f = \delta_k$ (Dirac-measure at point k) and $k \ge 2$, we have $\varphi_f(1) = \varphi_f(0)/2$, 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 $k \ge 0$ with $g_1[k] > 0$, we have

$$\frac{g_1[z]}{g_1[k]} \le \frac{g_2[z]}{g_2[k]} \text{ for all } z > k,$$

then $\bar{g}_1 \le \bar{g}_2$. Indeed, for $g_1[k] > 0$

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

and consequently

$$\sum_{z\ge k} g_1[z] \le \sum_{z\ge k} g_2[z].$$

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

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

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

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

$$\bar{g}_s = \frac{s^{r-1} + 2s^{r-2} + \cdots + r}{1 + s + \cdots + s^r}$$

is a decreasing function in s. Moreover, $\bar{g}_0 = r$ and $\bar{g}_1 = r/2$, and it follows for $0 \le s \le 1$

$$\frac{r}{2} \le \frac{r + (r-1)s + \cdots + s^{r-1}}{1 + s + \cdots + s^r} \le r. \qquad [1.11]$$

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

$$\begin{aligned} f'(1)(1-s) - (1 - f(s)) &= (1-s)\sum_{z=1}^{\infty} f[z] \sum_{k=0}^{z-1} (1 - s^k) \\ &= (1-s)^2 \sum_{z=1}^{\infty} f[z] \sum_{k=1}^{z-1}\sum_{j=0}^{k-1} s^j \\ &= (1-s)^2 \sum_{z=1}^{\infty} f[z]((z-1) + (z-2)s + \cdots + s^{z-2}). \end{aligned}$$

Therefore,

$$\begin{aligned} \varphi(s) &= \frac{f'(1)(1-s) - (1 - f(s))}{f'(1)(1-s)(1-f(s))} \\ &= \frac{\sum_{z=1}^{\infty} f[z]((z-1) + (z-2)s + \cdots + s^{z-2})}{\bar{f} \cdot \sum_{k=1}^{\infty} f[k](1 + s + \cdots + s^{k-1})}. \end{aligned}$$

From [1.11], it follows

$$\varphi(s) \le \frac{\psi(s)}{\bar{f}} \le 2\varphi(s) \qquad [1.12]$$

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 $g_s \in \mathcal{P}(\mathbb{N}_0)$, $0 \le s \le 1$, given by

$$g_s[z] := \frac{f[z+1](1 + s + \cdots + s^z)}{\sum_{k=0}^{\infty} f[k+1](1 + s + \cdots + s^k)}, \quad z \ge 0.$$

Then for $f[k+1] > 0$ and $z > k$, after some algebra,

$$\frac{g_s[z]}{g_s[k]} = \frac{f[z+1]}{f[k+1]} \prod_{v=1}^{z-k} \Big(1 + \frac{1}{s^{-1} + \cdots + s^{-k-v}}\Big),$$

which is an increasing function in s. Therefore,

$$\psi(s) = \bar{g}_s$$

is increasing in s. In combination with [1.12], we get

$$\varphi(s) \le \frac{\psi(s)}{\bar{f}} \le \frac{\psi(1)}{\bar{f}} \le 2\varphi(1),\ 2\varphi(s) \ge \frac{\psi(s)}{\bar{f}} \ge \frac{\psi(0)}{\bar{f}} \ge \varphi(0).$$

This gives the claim of the proposition. $\square$

PROOF (Proof of Theorem 1.1).– If $q = 1$, then $\mathbf{P}(\theta > n) \to 0$ 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 $\mathbf{P}(\theta > n) = \mathbf{P}(Z_n \neq 0) = 1 - f_{0,n}(0)$, such that it follows from Proposition 1.3

$$\frac{1}{\mathbf{P}(\theta > n)} = \frac{1}{\bar{f}_1 \cdots \bar{f}_n} + \sum_{k=1}^{n} \frac{\varphi_{f_k}(f_{k,n}(0))}{\bar{f}_1 \cdots \bar{f}_{k-1}}. \tag{1.13}$$

Moreover, we observe that $V2$ reads

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

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

$$\frac{\widetilde{f}_n}{4c} \le \varphi_{f_n}(s).$$

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

$$\begin{aligned}\frac{1}{\mathbf{P}(\theta > n)} &\ge \frac{1}{\bar{f}_1 \cdots \bar{f}_n} + \frac{1}{4c} \sum_{k=1}^{n} \frac{\widetilde{f}_k}{\bar{f}_1 \cdots \bar{f}_{k-1}} \\ &\ge \frac{1}{d} \frac{\mathbf{E}[Z_n(Z_n-1)] + \mathbf{E}[Z_n]}{\mathbf{E}[Z_n]^2} = \frac{1}{d}\Big(\frac{\mathbf{Var}[Z_n]}{\mathbf{E}[Z_n]^2} + 1\Big).\end{aligned}$$

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 $\mathcal{Z}$ in a varying environment. Under Assumption $V1$, an obvious choice is the process $\mathcal{W} = \{W_n, n \geq 0\}$, given by

$$W_n := \frac{Z_n}{\bar{f}_1 \cdots \bar{f}_n}, n \in \mathbb{N}_0,$$

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

$$\frac{Z_n}{\bar{f}_1 \cdots \bar{f}_n} \to W \text{ a.s. as } n \to \infty.$$

THEOREM 1.3.– *For a branching process $\mathcal{Z}$ 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 $\mathbf{E}[W] = 1$.

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

$$\sum_{k=1}^{\infty} \frac{\rho_k}{\bar{f}_1 \cdots \bar{f}_{k-1}} < \infty.$$

From [1.6], it follows that

$$\sup_n \frac{\mathbf{E}[Z_n^2]}{\mathbf{E}[Z_n]^2} = \sup_n \frac{\mathbf{Var}[Z_n]}{\mathbf{E}[Z_n]^2} + 1 < \infty.$$

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

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

$$Z_n \to Z_\infty \; a.s.$$

Moreover,

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

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 δ_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 $0 \le s < 1$. Then, given $n \in \mathbb{N}_0$, we define $l_n := n_k, m_n := n_{k+1}$ with $n_k \le n < n_{k+1}$, thus $l_n \le n < m_n$. We want to show that $f_{l_n,n}$ converges vaguely to δ_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

$$g(s) = g(h_1(h_2(s))), \quad 0 \le s < 1.$$

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 $n \to \infty$,

$$f_{0,n}(s) = f_{0,l_n}(f_{l_n,n}(s)) \to g(s), \quad 0 \le s < 1,$$

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 $0 \le s < 1$ 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 $(b_n, n \ge 0)$ of real numbers, such that $f_{0,n}(1/2) \le b_n \le 1$ and $b_n \downarrow g(1/2)$. We define the sequence $(a_n, n \ge 0)$ using the following equation:

$$f_{0,n}(a_n) = b_n.$$

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

$$\mathbf{E}[U_{n+1} \mid Z_0, \ldots, Z_n] = f_{n+1}(a_{n+1})^{Z_n} - f_{n+1}(0)^{Z_n} \leq a_n^{Z_n} - I\{Z_n = 0\} = U_n \text{ a.s.}$$

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 ∞.

Or $g = q\delta_0$. Then $g(1/2) = q$, implying that, for $n \to \infty$,

$$\mathbf{E}[U_n] = f_{0,n}(a_n) - f_{0,n}(0) = b_n - \mathbf{P}(Z_n = 0) \to g(1/2) - q = 0$$

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 ∞. In both cases, $Z_n \to Z_\infty$ a.s. for some random variable Z_∞.

iii) For the second statement, we use the representation $Z_n = \sum_{i=1}^{Z_{n-1}} Y_{i,n}$. Define the events $\mathcal{A}_{z,n} := \{\sum_{i=1}^{z} Y_{i,n} \neq z\}$. Then for $z \geq 1$

$$\mathbf{P}(\mathcal{A}_{z,n}) \geq 3^{-z}(1 - f_n[1]).$$

Indeed, if $f_n[1] \geq 1/3$, then

$$\begin{aligned}\mathbf{P}(\mathcal{A}_{z,n}) &\geq \mathbf{P}(Y_{1,n} \neq 1, Y_{2,n} = \cdots = Y_{z,n} = 1) \\ &\geq (1 - f_n[1]) f_n[1]^z \geq 3^{-z}(1 - f_n[1]),\end{aligned}$$

and if $f_n[1] \leq 1/3$, then either $\mathbf{P}(Y_{i,n} > 1) \geq 1/3$ or $\mathbf{P}(Y_{i,n} = 0) \geq 1/3$ implying

$$\begin{aligned}\mathbf{P}(A_{z,n}) &\geq \mathbf{P}(\min(Y_{1,n}, \ldots, Y_{z,n}) > 1) + \mathbf{P}(Y_{1,n} = \cdots = Y_{z,n} = 0) \\ &\geq 3^{-z}(1 - f_n[1]).\end{aligned}$$

Now assume $\sum_{n=1}^{\infty}(1 - f_n[1]) = \infty$. As, for fixed z, the events $\mathcal{A}_{z,n}$ 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 $z \geq 1$,

$$\mathbf{P}(Z_\infty = z) = \mathbf{P}(Z_n \neq z \text{ finitely often}) \leq \mathbf{P}(\mathcal{A}_{z,n} \text{ occurs finitely often}) = 0.$$

This implies that $\mathbf{P}(1 \leq Z_\infty < \infty) = 0$.