Lecture Notes in Statistics 219

Inés M. del Puerto · Miguel González
Cristina Gutiérrez · Rodrigo Martínez
Carmen Minuesa · Manuel Molina
Manuel Mota · Alfonso Ramos *Editors*

Branching Processes and Their Applications

Springer

Lecture Notes in Statistics

219

More information about this series at http://www.springer.com/series/694

Inés M. del Puerto • Miguel González •
Cristina Gutiérrez • Rodrigo Martínez •
Carmen Minuesa • Manuel Molina • Manuel Mota •
Alfonso Ramos
Editors

Branching Processes and Their Applications

Springer

Editors
Inés M. del Puerto
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Extremadura
Badajoz, Spain

Miguel González
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Extremadura
Badajoz, Spain

Cristina Gutiérrez
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Extremadura
Badajoz, Spain

Rodrigo Martínez
Departamento de Matemáticas
Centro Universitario de Plasencia
Universidad de Extremadura
Plasencia, Spain

Carmen Minuesa
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Extremadura
Badajoz, Spain

Manuel Molina
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Extremadura
Badajoz, Spain

Manuel Mota
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Extremadura
Badajoz, Spain

Alfonso Ramos
Departamento de Matemáticas
Facultad de Veterinaria
Universidad de Extremadura
Cáceres, Spain

ISSN 0930-0325 ISSN 2197-7186 (electronic)
Lecture Notes in Statistics
ISBN 978-3-319-31639-0 ISBN 978-3-319-31641-3 (eBook)
DOI 10.1007/978-3-319-31641-3

Library of Congress Control Number: 2016949394

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Foreword

Branching processes were proposed as a mathematical model for the survival of family names as well as the evolution of species over time (generations). Historically speaking, Francis Galton is credited with proposing the basic model for the survival of British peerage names in the 1870s. He posed it as a problem for solution in a London journal. It was solved, albeit incorrectly, by the Rev. Henry Watson, using generating functions of offspring distributions and their iterates. The correct surname extinction risk was found only 50 years later. Australian probabilists Eugene Seneta and Christopher Heyde, however, discovered that the French mathematician I.J. Bienaymé had derived the correct extinction probability almost 50 years before Galton formulated the problem and suggested that the basic model be named after all three, Bienaymé, Galton, and Watson.

The subject had major contributions from the Russian school of A.N. Kolmogorov and his many students in the 1940s, 1950s, and 1960s, in particular B.A. Sevastyanov. In the United States, the work and book of T.E. Harris led to a major growth of the field. Not only discrete time, single-type processes were considered but many extensions so as to include continuous-time and multiple types of individuals that were introduced. Also, in the early 1970s, Japanese and Canadian mathematicians introduced branching processes with movement over space as well. From the mid-1960s onwards, the subject exploded with important contributions from many parts of the world. New books were written and results appeared on both theory and applications.

Typically when a mathematical model for real-world phenomena is proposed and a theory is developed, it flourishes for a while and slowly fades away. This is not the case with branching processes. Over the last years, fundamental new contributions have continued to appear. Several international conferences have been held, in Asia, the United States, and in Europe, especially in Spain and Bulgaria.

Thus, the team of Dr. Inés M. del Puerto and her colleagues arranged a conference in 2009, 2012, and again in April 2015 at the University of Extremadura in Badajoz, Spain.

This volume presents papers from the latter event. As can be seen from the contents, the subject stays alive and well and open to fresh ideas. Participants came from all over the world.

Many thanks to the Extremadura team for arranging this conference and for their hospitality to the participants.

Bangalore, India
Melbourne, Australia
January 2016

Krishna B. Athreya
Peter Jagers

Preface

The Workshop on Branching Processes and their Applications (WBPA) was held during 7–10 April 2015 in Badajoz, Spain. This conference gave continuity to such important previous meetings organized with the aim to facilitate the exchange of research ideas in this field and related processes. The First World Congress on Branching Processes was held in Varna (Bulgaria) in 1993 to celebrate the first 150 years of branching processes. In the last ten years, it is worth mentioning those took place in Gothenburg (Sweden) in 2005, in Luminy (France) in 2007 and 2011, in Beijing (China) in 2011, and in Pomorie (Bulgaria) in 2012 and 2014.

The meeting was the third in the series of WBPAs promoted and organized by the Branching Processes Research Group belonging to the Department of Mathematics of the University of Extremadura, Spain. It was attended by 36 invited participants from 13 countries from all over the world. The papers presented at the workshop maintained a healthy balance between the theoretical and practical aspects of branching process theory, showing it to be an area of active and interesting research. The program and abstracts are available on the conference website (http://branching.unex.es/wbpa15/index.htm).

The Proceedings consists of 19 papers. All of them have been thoroughly reviewed. The topics covered by the workshop have been classified into the following areas (articles are sorted in alphabetical order by author inside each area):

1. Coalescent branching processes
2. Branching random walks
3. Population growth models in varying and random environments
4. Size-density-resource-dependent branching models
5. Age-dependent branching models
6. Special branching models
7. Applications in epidemiology
8. Applications in biology and genetics

The first part deals with coalescent branching processes. K.B. Athreya studies, for Galton-Watson branching trees, the exact distribution of the generation number of the last common ancestor of two individuals randomly chosen in generation n (provided there exist at least two individuals in this generation). He also studies its asymptotic behavior when n goes to infinity. He considers four cases: explosive (mean infinity), supercritical, critical, and subcritical. R.C. Griffiths discusses multi-type Λ-coalescent processes which arise naturally from Λ-Fleming-Viot processes as dual processes back in time. He considers that mutation and selection may be modeled as happening at random in the population or in families at birth.

Branching random walks are studied in Part II. E. Antonenko and E. Yarovaya consider a continuous-time branching random walk on a multidimensional lattice with finite variance of jumps and a finite set of the particle generation centers, i.e., branching sources. They focus on the study of the evolutionary operator for the mean number of particles both at an arbitrary point and on the entire lattice. W. Hong and H. Wang survey branching structures within random walks. As an example of application, the stationary distribution of a birth-and-death process with bounded jumps is calculated.

Part III is devoted to population growth models in varying and random environments. M. Ispány analyzes strongly critical branching processes with immigration in varying environment, obtaining a diffusion approximation when either the offspring or the immigration variances are strictly positive, and a fluctuation limit theorem in the case of asymptotically vanishing offspring variances. V. Vatutin presents a survey of some recent results on the asymptotics of the survival probability, limit theorems conditioned on survival or attaining a high level of single-type subcritical branching processes in independent and identically distributed random environments.

Part IV deals with size-density-resource-dependent branching models. F.T. Bruss uses resource-dependent branching processes to model the evolution of societies. He analyzes the implications of his (and Duerinckx) *Theorem of Envelopment of Societies*, which displays that every human population (modeled by this kind of process) will be bound, in the long run, to fluctuate between extreme society forms, for whatever way it distributes resources in different generations. He also attracts interest to resource-dependent branching processes as a way to model complicated growth processes. P. Jagers and F.C. Klebaner consider a general branching process where, at any moment, the individual birth-and-death intensities as well as the splitting distribution can be influenced by the size and composition of the whole population. They present special cases of this process as models for populations evolving in habitats with finite carrying capacity and, in the multitype case, for two-sex populations.

Part V is devoted to age-dependent branching models. O. Hyrien et al. deal with a Sevastyanov age-dependent branching process with immigration at times given by a time-nonhomogeneous Poisson process. This model can be used to describe the dynamics of cell populations arising from differentiating stem cells. They provide results on the asymptotic behavior of these processes for various classes of immigration rates in the supercritical case. P. Trayanov uses the Crump-Mode-Jagers branching process to model the evolution of a population. He presents a numerical approach for projecting a population age structure and solving the corresponding renewal equations that give us the expected future size of the population.

Part VI deals with Special Branching Processes. M. González et al. focus their attention on controlled branching processes. They develop the inferential theory for these models from a Bayesian outlook, considering a nonparametric setting for the offspring law and a parametric one for the control distributions. They use Markov chain Monte Carlo methods when dealing with incomplete data schemes. G. Kersting studies criteria for the recurrence and transience of a general multivariate stochastic process fulfilling a nonlinear stochastic difference equation similar to the one verified by near-critical branching processes. U. Roesler gives an overview of research on the weighted branching process. This model has many applications in genetics, computer science, and algorithms and includes as particular cases, for example, Biggins branching random walks, Mandelbrot cascades, and the Quicksort process. S. Sagitov and A. Lindo consider Galton-Watson branching processes whose reproduction laws are an extension of the two-parameter linear-fractional family to a four-parameter family. They study this especial class of branching processes showing that in some explosive cases the time to explosion can be approximated by the Gumbel distribution.

Part VII comprises some applications of the branching processes theory in epidemiology. F. Ball et al. are concerned with the use of vaccination schemes to control an epidemic in terms of the total number of infected individuals. They establish monotonicity and continuity properties of total progeny of Crump-Mode-Jagers branching processes depending on vaccination level and derive, from them, optimal vaccination polices based on the mean and quantiles of the total number of infected individuals. F. Ball and L. Shaw consider the problem of estimating the within-household infection rate for a Markov susceptible-infective-recovered epidemic among a population that is partitioned into households, from observation of the early, exponentially growing phase of an epidemic. They use the asymptotic theory of continuous-time Markov branching processes to approach this problem and show, by simulations, that the proposed method is feasible for realistic population sizes.

Part VIII focuses on applications in biology and genetics. M. González et al. introduce a multitype two-sex branching process for describing the evolution of the number of carries of the two alleles—recessive and dominant—of a gene linked to X-chromosome. They consider recessive pernicious alleles (responsible for a disorder) and study their fate in the population. Moreover, they investigate conditions for the fixation of the dominant alleles as well as for the coexistence of both types of alleles. M. Molina et al. deal with the extinction/survival of populations modeled by two-sex branching processes where mating and reproduction are affected by the number of females and males in the population. M. Slavtchova-Bojkova considers the problem of modeling the dynamics of the number of different types of cells, which due to a small reproductive ratio are fated to become extinct but can mutate during the reproduction process and become a new type of cells that may escape extinction. This problem is usual in the cancer setting. She uses a continuous-time age-dependent branching process to model this situation and derives the numbers of mutant cells of the escape type and their moments.

The organizers greatly appreciate the major response from the participants to submit contributions to the Proceedings. We also like to acknowledge the reviewers who helped improve the papers. They would also like to thank all the people who actively participated in organizing the workshop and those entities which provided financial and scientific support. The main funding came from the Junta de Extremadura, Spanish Ministry of Economy and Competitiveness (Ministerio de Economía y Competitividad), the University of Extremadura itself, and the local administration. The Sociedad de Estadística e Investigación Operativa (SEIO) and the Instituto de Computación Científica Avanzada de Extremadura (ICCAEx) supported the meeting scientifically.

Also, our very special thanks go to Professors Krishna B. Athreya and Peter Jagers for accepting the task of writing the foreword of this book. Finally, many thanks to the editorial office of Springer Verlag for making it possible for this project to see the light.

Badajoz, Spain
January 2016

Inés M. del Puerto
On behalf of the Editors

Contents

Part I Coalescent Branching Processes

Part II Branching Random Walks

Part III Population Growth Models in Varying and Random Environments

Part IV Size/Density/Resource-Dependent Branching Models

Part V Age-Dependent Branching Models

Part VI Special Branching Models

Part VII Applications in Epidemiology

Part VIII Applications in Biology and Genetics

List of Contributors

Ekaterina Antonenko Department of Probability Theory, Lomonosov Moscow State University, Moscow, Russia

Krishna B. Athreya Iowa State University, Ames, IA, USA

Frank Ball School of Mathematical Sciences, University of Nottingham, Nottingham, UK

F. Thomas Bruss Département de Mathèmatique, Université Libre de Bruxelles, Bruxelles, Belgium

Inés M. del Puerto Department of Mathematics, University of Extremadura, Avda. Elvas, Badajoz, Spain

Miguel González Department of Mathematics, University of Extremadura, Avda. Elvas, Badajoz, Spain

Robert C. Griffiths University of Oxford, Oxford, UK

Cristina Gutiérrez Department of Mathematics, University of Extremadura, Avda. Elvas, Badajoz, Spain

Wenming Hong School of Mathematical Sciences & Key Laboratory of Mathematics and Complex Systems, Beijing Normal University, Beijing, China

Ollivier Hyrien Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA

Márton Ispány Faculty of Informatics, University of Debrecen, Debrecen, Hungary

Peter Jagers Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, Gothenburg, Sweden

Götz Kersting Institut für Mathematik, Goethe Universität, Frankfurt am Main, Frankfurt, Germany

Fima C. Klebaner School of Mathematical Sciences, Monash University, Clayton, VIC, Australia

Alexey Lindo Chalmers University and University of Gothenburg, Gothenburg, Sweden

Rodrigo Martínez Centro Universitario de Plasencia, Department of Mathematics, University of Extremadura, Plasencia, Spain

Carmen Minuesa Department of Mathematics, University of Extremadura, Avda. Elvas, Badajoz, Spain

Kosto V. Mitov Faculty of Aviation, National Military University "Vasil Levski", Pleven, Bulgaria

Manuel Molina Faculty of Sciences, Department of Mathematics, University of Extremadura, Badajoz, Spain

Manuel Mota Faculty of Sciences, Department of Mathematics, University of Extremadura, Badajoz, Spain

Alfonso Ramos Facultad de Veterinaria, Department of Mathematics, University of Extremadura, Cáceres, Spain

Uwe Roesler Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Serik Sagitov Chalmers University and University of Gothenburg, Gothenburg, Sweden

Laurence Shaw School of Mathematical Sciences, University of Nottingham, Nottingham, UK

Maroussia Slavtchova-Bojkova Department of Probability, Operations Research and Statistics, Faculty of Mathematics and Informatics, Sofia University, Sofia, Bulgaria

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

Plamen Trayanov Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

Vladimir Vatutin Steklov Mathematical Institute, Moscow, Russia

Huaming Wang Department of Statistics, Anhui Normal University, Wuhu, China

Nikolay M. Yanev Department of Probability and Statistics, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

Elena Yarovaya Department of Probability Theory, Lomonosov Moscow State University, Moscow, Russia

Part I
Coalescent Branching Processes

Chapter 1
Coalescence in Branching Processes

Krishna B. Athreya

Mathematics Subject Classification (2000): 60J80, 60F10

1.1 Introduction: The Problem of Coalescence in Trees

Let $\mathcal{T}$ be a rooted tree. Let $\{v_{n1}, v_{n2}, \ldots, v_{nZ_n}\}$ be the set of vertices at the nth level. Pick two of the v_{ni}'s by SRSWOR (simple random sampling without replacement) (assuming $Z_n \geq 2$) and trace their lines of descent back in time till they meet for the first time. Call that generation X_n.

The problems are:

a) Find the distribution of X_n.
b) Study its limit as $n \to \infty$.

X_n is called the *coalescence time*.

c) Do the same with choosing k vertices out of Z_n.
d) Do the same with choosing all Z_n vertices out of Z_n.

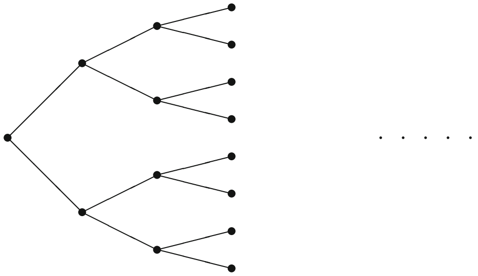
Clearly, the answers depend on how $\mathcal{T}$ is generated.

K.B. Athreya (✉)
Iowa state University, Ames, IA 50011, USA
e-mail: kbathreya@gmail.com

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_1

1.2 The Binary Tree Case

Consider a binary tree $\mathcal{T}$ starting with one vertex. The tree looks like



At level n, there are 2^n vertices, $n = 0, 1, 2, \ldots$. Pick two vertices at level n by SRSWOR. Trace their lines back till they meet. Call that generation X_n. Then, for $k = 1, 2, \ldots, n$, the probability that $X_n < k$ is,

$$P(X_n < k) = \frac{\binom{2^k}{2} 2^{n-k} 2^{n-k}}{\binom{2^n}{2}} = \frac{2^k(2^k - 1)2^{n-k}2^{n-k}}{2^n(2^n - 1)} = \frac{1 - 2^{-k}}{1 - 2^{-n}}.$$

So, $\lim_{n\to\infty} P(X_n < k) = 1 - 2^{-k}$, $k = 1, 2, \ldots$. Thus, X_n converges in distribution to geometric distribution with parameter $\frac{1}{2}$. Similar result is true for any regular $b-$nary tree, $b \geq 2$. This suggests that a similar result must be true for a growing Galton–Watson tree.

1.3 Galton–Watson Trees

1.3.1 Definition and the Problem

Let $\{p_j\}_{j\geq 0}$ be a probability distribution on $\mathbb{N}^+ \equiv \{0, 1, 2, \ldots\}$, $\{\xi_{n,i} : i \geq 1, n \geq 0\}$ be i.i.d. with common probability distribution $\{p_j\}_{j\geq 0}$, Z_0 be a positive integer, and for $n \geq 0$,

$$Z_{n+1} = \begin{cases} \sum_{i=1}^{Z_n} \xi_{n,i}, \ n \geq 0 & \text{if } Z_n > 0, \\ 0 & \text{if } Z_n = 0. \end{cases}$$

Then $\{Z_n\}_{n\geq 0}$ is called a Galton-Watson branching process with initial population Z_0 and offspring distribution $\{p_j\}_{j\geq 0}$, and $\xi_{n,i}$ is the number of offspring of the ith individual of the nth generation.

Now, every individual in the nth generation, $n \geq 1$, can be identified by a finite string

$$u_n \equiv (i_0, i_1, i_2, \ldots, i_n)$$

meaning that this individual is the i_nth offspring of the $u_{n-1} \equiv (i_0, i_1, \ldots, i_{n-1})$ and $u_0 = i_0$ is the number associated with the i_0th member of the 0th generation.

Let $A_{n,2} \equiv \{Z_n \geq 2\}$ and $B_n \equiv \{Z_n > 1\}$ be events defined on the space of family trees.

Consider the following questions:

3.1 a) Conditioned on $A_{n,2}$, pick two individuals in the nth generation by SRSWOR and trace their lines back till they meet. Call that generation $X_{n,2}$.
What is the distribution of $X_{n,2}$?
What happens to it as $n \to \infty$?

3.1 b) Do the same thing with k choices ($2 \leq k < \infty$) by SRSWOR from the nth generation. Call the coalescence time $X_{n,k}$. Ask the same questions.

3.1 c) Do the same thing for the whole population. Call the coalescence time Y_n. Ask the same questions.

Note that there are two sources of randomness here: one, that of the tree $\mathcal{T}$ being random and two, the sampling of vertices at the nth stage.

1.3.2 Basic Results for Galton-Watson Trees

We now state some known results on branching processes.

Theorem 1.1 (Supercritical Case) *Let* $p_0 = 0$, $1 < m \equiv \sum_{j=1}^{\infty} jp_j < \infty$. *Then*

a) $P(Z_n \to \infty \mid Z_0 > 0) = 1$.

b) $\left\{W_n \equiv \dfrac{Z_n}{m^n} : n \geq 0\right\}$ *is a nonnegative martingale and hence*

$$\lim_{n\to\infty} W_n \equiv W \text{ exists w.p.1.}$$

c) $\sum_{j=1}^{\infty}(j \log j)p_j < \infty$ *implies* $E(W \mid Z_0 = 1) = 1$.

Further, in this case W *has an absolutely continuous distribution on* $(0,\infty)$ *with a positive density. Also* $\sum_{j=1}^{\infty}(j\log j)p_j = \infty$ *implies* $P(W = 0 \mid Z_0 = 1) = 1$.

d) *There always exists a sequence* $\{C_n\}_{n\geq 1}$ *such that*

$$\frac{C_{n+1}}{C_n} \to m \quad \text{and} \quad \frac{Z_n}{C_n} \to W \text{ w.p.1}, \quad \text{as } n \to \infty,$$

and $P(0 < W < \infty) = 1$.

For proofs of a) to d) above see Athreya and Ney [7].

Theorem 1.2 (Critical Case) *Let* $m = 1$, $p_j \neq 1$ *for any* $j \geq 1$ *and* $0 < \sigma^2 \equiv \sum_{j=1}^{\infty} j^2 p_j - 1 < \infty$. *Then*

a) $P(Z_n \to 0 \mid Z_0 > 0) = 1$.

b) $nP(Z_n > 0) \to \dfrac{\sigma^2}{2} \quad \text{as } n \to \infty$.

c) $P\left(\dfrac{Z_n}{n} > x \,\middle|\, Z_n > 0\right) \to e^{-\frac{2}{\sigma^2}x}, \quad \text{as } n \to \infty,\ 0 < x < \infty$.

d) *For* $1 \leq k \leq n$, *let*

$$V_{n,k} \equiv \left\{ \frac{Z^{(k)}_{n-k,i}}{n-k} I(Z^{(k)}_{n-k,i} > 0) : 1 \leq i \leq Z_k \right\}$$

on the event $\{Z_k > 0\}$, *where* $\{Z^{(k)}_{j,i} : j \geq 0\}$ *is the Galton–Watson process initiated by the ith individual in the kth generation. Let* $k \to \infty$, $n \to \infty$ *such that* $\frac{k}{n} \to u$, $0 < u < 1$. *Then the sequence of point processes* $\{V_{n,k}\}_{n\geq 1}$ *conditioned on* $\{Z_n \geq 1\}$ *converges weakly to the point process*

$$V \equiv \{\eta_j : j = 1, 2, \ldots, N_u\},$$

where $\{\eta_j\}_{j\geq 1}$ *are i.i.d.* $\exp(2/\sigma^2)$, N_u *is Geom*(u), *i.e.,* $P(N_u = k) = (1-u)u^{k-1}$, $k \geq 1$ *and* $\{\eta_j\}_{j\geq 1}$ *and* N_u *are independent.*

For proofs of a), b) and c) above see Athreya and Ney [7]. For d) see Athreya [4].

Theorem 1.3 (Subcritical Case) *Let* $0 < m < 1$. *Then*

a) *For* $j \geq 1$, $\lim\limits_{n\to\infty} P(Z_n = j \mid Z_n > 0) \equiv b_j$ *exists,* $\sum\limits_{j=1}^{\infty} b_j = 1$ *and* $B(s) \equiv \sum\limits_{j=1}^{\infty} b_j s^j$, $0 \leq s \leq 1$, *is the unique solution of the functional equation*

$$B(f(s)) = mB(s) + (1-s) \quad ,0 \leq s \leq 1,$$

where $f(s) \equiv \sum_{j=0}^{\infty} p_j s^j$, *in the class of all probability generating functions vanishing at* 0.

b) $\sum_{j=1}^{\infty} jb_j < \infty$ *iff* $\sum_{j=1}^{\infty} (j \log j) p_j < \infty$.

c) $\lim_{n\to\infty} \frac{P(Z_n > 0 \mid Z_0 = 1)}{m^n} = \frac{1}{\sum_{j=1}^{\infty} jb_j}$.

Let Z_0 *be a random variable. Then*

d) If $EZ_0 < \infty$, *then*

$$\lim_{n\to\infty} P(Z_n = j \mid Z_n > 0) = b_j \quad , j \geq 1$$

and if, in addition, $\sum_{j=1}^{\infty} (j \log j) p_j < \infty$ *then*

$$\sum_{j=1}^{\infty} jb_j < \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{P(Z_n > 0)}{m^n} = \frac{EZ_0}{\sum_{j=1}^{\infty} jb_j}.$$

For proofs of a) to d) above see Athreya and Ney [7].

1.4 Coalescence Results for Galton-Watson Trees

We now present some results on coalescence for Galton-Watson trees.

Theorem 1.4 (Supercritical Case) *Let* $p_0 = 0$, $1 < m < \infty$. *Then, for almost all trees* $\mathcal{T}$,

i) for all $1 \leq k < \infty$,

$$\lim_{n\to\infty} P(X_{n,2} < k \mid \mathcal{T}) \equiv \pi_{k,2}(\mathcal{T}) \quad \text{exists}$$

and $\pi_{k,2}(\mathcal{T}) \uparrow 1$ *as* $k \uparrow \infty$.

ii) for all $j \geq 2$ *and for all* $\forall 1 \leq k < \infty$,

$$\lim_{n\to\infty} P(X_{n,j} < k \mid \mathcal{T}) \equiv \pi_{k,j}(\mathcal{T}) \quad \text{exists}$$

and $\pi_{k,j}(\mathcal{T}) \uparrow 1$ *as* $k \uparrow \infty$.

iii) Let $p_1 > 0$. *Then, for almost all trees* $\mathcal{T}$,

$$Y_n \to N(\mathcal{T})$$

where $N(\mathcal{T}) = \max\{j \geq 1 : Z_j = 1\}$. *Also,*

$$\lim_{n\to\infty} P(Y_n = k) = (1-p_1)p_1^k \quad ,k \geq 0.$$

Theorem 1.5 (Critical Case) *Let* $m = 1$, $p_1 < 1$ *and* $\sigma^2 = \sum_{j=1}^{\infty} j^2 p_j - 1 < \infty$, *Then, for* $0 < u < 1$,

i) $\lim_{n\to\infty} P\left(\frac{X_{n,2}}{n} < u \,\middle|\, Z_n \geq 2\right) \equiv H_2(u)$ *exists and for* $0 < u < 1$,

$$H_2(u) \equiv 1 - E\varphi(N_u)$$

where N_u *is a geometric random variable with distribution*

$$P(N_u = k) = (1-u)u^{k-1} \quad ,k \geq 1,$$

and for $j \geq 1$,

$$\varphi(j) \equiv E\left(\frac{\sum_{i=1}^{j} \eta_i^2}{(\sum_{i=1}^{j} \eta_i)^2}\right)$$

where $\{\eta_i\}_{i\geq 1}$ *are i.i.d. exponential r.v. with* $E\eta_1 = 1$.

Further, $H_2(\cdot)$ *is absolutely continuous on* $[0,1]$, $H_2(0+) = 0$, *and* $H_2(1-) = 1$.

ii) For $0 < u < 1$, $1 < k < \infty$,

$$\lim_{n\to\infty} P\left(\frac{X_{n,k}}{n} < u \,\middle|\, Z_n \geq k\right) \equiv H_k(u) \quad \text{exists}$$

and $H_k(\cdot)$ *is absolutely continuous function with* $H_k(0+) = 0$ *and* $H_k(1-) = 1$.

iii) For $0 < u < 1$, $\lim_{n\to\infty} P\left(\frac{Y_n}{n} < u \,\middle|\, Z_n \geq 1\right) = u$.

Theorem 1.6 (Subcritical Case) *Let* $0 < m < 1$. *Then*

i) For $k \geq 1$, $\lim_{n\to\infty} P(n - X_n > k \mid Z_n \geq 2) = \frac{E\phi_k(Y)}{E\psi_k(Y)} \equiv \pi_k$, *say, where*

$$\phi_k(j) = E\left(\frac{\sum\limits_{i_1 \neq i_2 = 1}^{j} Z_{k,i_1} Z_{k,i_2}}{\left(\sum\limits_{i=1}^{j} Z_{k,i}\right)\left(\sum\limits_{i=1}^{j} Z_{k,i} - 1\right)} I\left(\sum_{i=1}^{j} Z_{k,i} \geq 1\right)\right)$$

and

$$\psi_k(j) = P\left(\sum_{i=1}^{j} Z_{k,i} \geq 2\right)$$

where $\{Z_{r,i} : r \geq 0\}$, $i = 1, 2, \ldots$, *are i.i.d. copies of a Galton-Watson branching process* $\{Z_r : r \geq 0\}$ *with* $Z_0 = 1$ *and the given offspring distribution* $\{p_j\}_{j\geq 0}$ *and Y is a random variable with distribution* $\{b_j\}_{j\geq 1}$ *where*

$$b_j \equiv \lim_{n\to\infty} P(Z_n = j \mid Z_n > 0, Z_0 = 1) \quad \textit{which exists.}$$

Further, if $\sum_{j=1}^{\infty} j \log j p_j < \infty$, *then* $\lim_{k\uparrow\infty} \pi_k = 0$ *and hence* $n - X_n$ *conditioned on* $Z_n \geq 2$ *converges to a proper distribution on* $\{1, 2, \ldots\}$.

ii) For $k \geq 1$, $\lim_{n\to\infty} P(n - Y_n > k \mid Z_n \geq 1) \equiv \tilde{\pi}_k$ *exists and equals*

$$E\left(\frac{1 - q_k^Y}{m^k}\right) - E\left(\frac{Y q^{k-1}(1 - q_k)}{m^k}\right)$$

where Y is a random variable with distribution

$$P(Y = j) = b_j = \lim_{n\to\infty} P(Z_n = j \mid Z_n > 0, Z_0 = 1)$$

and $q_k = P(Z_k = 0 \mid Z_0 = 1)$.

Further, if $\sum_{j=1}^{\infty} j \log j p_j < \infty$, *then* $\lim_{k\to\infty} \tilde{\pi}_k = 0$. *That is,* $n - Y_n$ *conditioned on* $\{Z_n > 0\}$ *converges in distribution as* $n \to \infty$ *to a proper distribution on* $\{1, 2, \ldots\}$.

For proofs of Theorems 1.4–1.6 see Athreya [3, 4]. In what follows $\xrightarrow{d}$ means convergence in distribution.

Theorem 1.7 (Explosive Case) *Let* $p_0 = 0$*,* $m = \infty$*, and for some* $0 < \alpha < 1$*, and a function* $L : (1,\infty) \to (0,\infty)$ *slowly varying at* ∞*, i.e., for all* $0 < c < \infty$*,*

$$\frac{L(cx)}{L(x)} \to 1 \quad \text{as } x \to \infty.$$

Let

$$\frac{\sum\limits_{j>x} p_j}{x^{\alpha} L(x)} \to 1 \quad \text{as } x \to \infty.$$

Then

i) $\alpha^n \log Z_n \to \eta$ *w.p.1 and* $P(0 < \eta < \infty) = 1$ *and* η *has a continuous distribution.*

ii) Let $\{Z_n^{(1)}\}_{n\geq 1}$ *and* $\{Z_n^{(2)}\}_{n\geq 1}$ *be two i.i.d. copies of a Galton–Watson branching process with* $\{p_j\}_{j\geq 1}$ *satisfying the above hypotheses. Then, w.p.1*

$$\frac{Z_n^{(1)}}{Z_n^{(2)}} \to \begin{cases} 0 & \text{with prob. } \frac{1}{2} \\ \infty & \text{with prob. } \frac{1}{2} \end{cases}$$

iii) For almost all trees $\mathcal{T}$ *and* $k = 1, 2, \ldots$*, as* $n \to \infty$*,*

$$P(X_{n,2} < k \mid \mathcal{T}) \to 0$$

and

$$P(n - X_{n,2} < k) \to \pi_2(k) \quad \text{exists}$$

and $\pi_2(k) \uparrow 1$ *as* $k \uparrow \infty$*.*

iv) For any $1 < j < \infty$ *and* $k = 1, 2, \ldots$

$$P(X_{n,j} < k \mid \mathcal{T}) \to 0 \quad \text{as } n \to \infty$$

and $P(n - X_{n,j} < k) \to \pi_j(k)$ *exists and* $\pi_j(k) \uparrow 1$ *as* $k \uparrow \infty$*.*

v) $Y_n \xrightarrow{d} N(\mathcal{T}) \equiv \max\{j : Z_j = 1\} < \infty$ *and*

$$P(Y_n = k) \to (1 - p_1)p_1^{k-1} \quad , k \geq 1.$$

The proof of Theorem 1.7 needs the following results.

Proposition 1.1 *Let* $\{Z_n\}_{n\geq 0}$ *be a Galton–Watson branching process with offspring distribution* $\{p_j\}_{j\geq 0} \in D(\alpha)$, *(domain of attraction of a stable law of order* α*),* $0 < \alpha < 1$, *and* $Z_0 = 1$. *Then,*

$$Z_k \in D(\alpha^k) \quad \textit{for all} \;\; 1 \leq k < \infty.$$

Proposition 1.2 *Let* $\{X_i\}_{i\geq 1}$ *be i.i.d. random variables such that* $P(0 < X_1 < \infty) = 1$ *and* $X_1 \in D(\alpha)$, $0 < \alpha < 1$. *Then*

a)

$$\frac{\sum_{i=1}^{n} X_i^2}{\left(\sum_{i=1}^{n} X_i\right)^2} \xrightarrow{d} Y_\alpha$$

where Y_α *is a continuous r.v. with* $P(0 < Y_\alpha < 1) = 1$.

b) $EY_\alpha \uparrow 1$ *as* $\alpha \downarrow 0$.

c) For any $j = 2, 3, \ldots$,

$$\frac{\sum_{i=1}^{n} X_i^j}{\left(\sum_{i=1}^{n} X_i\right)^j} \xrightarrow{d} Y_{\alpha,j}$$

and $EY_{\alpha,j} \uparrow 1$ *as* $\alpha \downarrow 0$.

For proofs of the results in Theorem 1.7 and Propositions 1.1 and 1.2, see Athreya [3, 4]. Here is the basic calculation:

$$\begin{aligned} P(X_n \geq k \mid \mathcal{T}) &= \frac{\sum_{i=1}^{Z_k} \binom{Z_{n-k,i}^{(k)}}{2}}{\binom{Z_n}{2}} \\ &= \frac{\sum_{i=1}^{Z_k} Z_{n-k,i}^{(k)}\big(Z_{n-k,i}^{(k)} - 1\big)}{\left(\sum_{i=1}^{Z_k} Z_{n-k,i}^{(k)}\right)\left(\sum_{i=1}^{Z_k} Z_{n-k,i}^{(k)} - 1\right)} \end{aligned}$$

a) $1 < m < \infty$

Fix k, by Seneta–Heyde, there exists a sequence $\{C_n\}_{n\geq 0}$ of constants such that

$$\frac{Z_{n-k,i}^{(k)}}{C_{n-k}} \to W_{k,i} \quad \text{w.p.1}$$

and $P(0 < W_{k,i} < \infty) = 1$. So,

$$P(X_n \geq k \mid \mathcal{T}) \to \frac{\sum_{i=1}^{Z_k} W_{k,i}^2}{\left(\sum_{i=1}^{Z_k} W_{k,i}\right)^2} \quad \text{w.p.1}$$

and this can be shown to converge to 0 as $k \to \infty$ (see [3]).

b) $m = \infty$, $\{p_j\} \in D(\alpha)$, $0 < \alpha < 1$.

$$\begin{aligned} P(n - X_n \leq k) &= P(X_n \geq n - k) \\ &= E\left(\frac{\sum_{i=1}^{Z_{n-k}} Z_{k,i}^{(n-k)}\big(Z_{k,i}^{(n-k)} - 1\big)}{Z_n(Z_n - 1)}\right). \end{aligned}$$

This can be shown to converge to $\pi(k)$ and $\pi(k) \uparrow 1$ as $k \uparrow \infty$ (see [3]).

c) Similar argument for $m = 1$ and $0 < m < 1$ (need point process result for $m = 1$ and Theorem 1.3, a) for $0 < m < 1$).

Summary:

a) $1 < m < \infty$:
  $X_{n,2} \xrightarrow{d}$ a proper distribution on $\mathbb{N}^+$

b) $m = \infty$, $\{p_j\}_{j\geq 0} \in D(\alpha)$, $0 < \alpha < 1$:
  $n - X_{n,2} \xrightarrow{d}$ a proper distribution on $\mathbb{N}^+$

c) $m = 1, \sigma^2 < \infty$:
  i) $\left.\frac{X_{n,2}}{n}\,\right|\, Z_n \geq 2 \xrightarrow{d}$ absolutely continuous distribution on $[0, 1]$
  ii) $\left.\frac{Y_n}{n}\,\right|\, Z_n \geq 1 \xrightarrow{d}$ uniform distribution on $[0, 1]$

d) $0 < m < 1$:
  $\left.\big(n - X_{n,2}\big)\,\right|\, Z_n \geq 2 \xrightarrow{d}$ a proper distribution on $\{1, 2, \ldots\}$

i.e.,

a) $1 < m < \infty$: coalescence is near the beginning of the tree.
b) $m = \infty$, $\{p_j\}_{j\geq 0} \in D(\alpha)$, $0 < \alpha < 1$: coalescence is near the present.
c) $m = 1, \sigma^2 < \infty$: $X_{n,2}$ is of order n.
d) $0 < m < 1$: $X_{n,2}$ is near the present.

1.5 Application of Coalescence Results to Branching Random Walks

Let $\mathcal{T}$ be a Galton-Watson tree with $Z_0 = 1$ and offspring distribution $\{p_j\}_{j\geq 0}$. Impose on this tree $\mathcal{T}$ the following movement structure:

If an individual is at x in $\mathbb{R}$ and has k children then these k children move to $x + X_{k,j}$, $j = 1, 2, \ldots, k$, where $X_k \equiv (X_{k,1}, X_{k,2}, \ldots, X_{k,k})$ has a joint distribution $\pi_k(\cdot)$ on $\mathbb{R}^k$.

Also, assume that the random vector X_k is stochastically independent of the history up to that generation as well as the movement of the other individuals of that generation.

Let Z_n be the number of individuals in the nth generation and $\zeta_n \equiv \{x_{n,i} : 1 \leq i \leq Z_n\}$ be the positions of the Z_n individuals of the nth generation.

A problem of interest is what happens to the point process ζ_n as $n \to \infty$.

In what follows, let $Z_n(x)$ be the number of $X_{n,i}$ less than or equal to x, for $1 \leq i \leq Z_n$, i.e., $Z_n(x) \equiv \sum_{i=1}^{Z_n} I(X_{n,i} \leq x)$.

1.5.1 Supercritical Case

Theorem 1.8 (Athreya [2]) *Let* $p_0 = 0$*,* $1 < m < \infty$ *and* π_k *be such that* $\{X_{k,i} : i = 1, 2, \ldots, k\}_{k\geq 1}$ *are identically distributed.*

a) Let $EX_{k,1} = 0$ *and* $EX_{k,1}^2 = \sigma^2 < \infty$*. Then, for all* $y \in \mathbb{R}$*,*

$$\frac{Z_n(\sqrt{n}\sigma y)}{Z_n} \to \Phi(y) \quad \textit{(the standard } N(0,1) \textit{ cdf)}$$

in mean square.

b) If $X_{k,1} \in D(\alpha)$*,* $0 < \alpha \leq 2$*, then there exist sequences* $\{a_n\}_{n\geq 1}$ *and* $\{b_n\}_{n\geq 1}$ *such that*

$$\frac{Z_n(a_n + b_n y)}{Z_n} \to G_\alpha(y) \quad \textit{in mean square,}$$

where $G_\alpha(\cdot)$ *is a standard stable law cdf (of order* α*).*

c) In a), if Y_n *is the position of a randomly chosen individual from the nth generation with uniform distribution, then, for all* $y \in \mathbb{R}$*,*

$$P(Y_n \leq \sqrt{n}\sigma y) \to \Phi(y)$$

and similarly for b).

The proof depends on the fact when $p_0 = 0$ and $1 < m < \infty$, the coalescence time $X_{n,2}$ is way back in time and so the positions of two randomly chosen individuals in the nth generation are essentially independent and have the marginal distribution of a random walk at step n. For full proof see Athreya [2].

1.5.2 Explosive Case

Theorem 1.9 *Let $m = \infty$, $\{p_j\}_{j\geq 0} \in D(\alpha)$, $0 < \alpha < 1$. Let $\{X_{k,i} : 1 \leq i \leq k\}_{k\geq 1}$ be identically distributed. Let $EX_{k,1} = 0$ and $EX_{k,1}^2 = \sigma^2 < \infty$. Then, for each fixed y in $\mathbf{R}$,*

$$\frac{Z_n(\sqrt{n}\sigma y)}{Z_n} \xrightarrow{d} \delta_y$$

where δ_y is Bernoulli($\Phi(y)$), i.e.

$$\delta_y = \begin{cases} 1, & \textit{with prob. } \Phi(y) \\ 0, & \textit{with prob. } 1 - \Phi(y) \end{cases}$$

The proof is based on the fact that

$$E\left(\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}\right)^k \to \Phi(y) \quad \text{for } k = 1, 2.$$

This, in turn, is due to the fact that $X_{n,2}$, the coalescence time for any two individuals chosen at random from the nth generation is such that $n - X_{n,k}$ converges to a proper distribution (Theorem 1.7) and hence their positions differ only by an amount that converges in distribution.

This can be strengthened to joint convergence of the finite dimensional distribution of

$$\left\{\frac{Z_n(\sqrt{n}\sigma y)}{Z_n}, \quad i = 1, 2, \ldots, k, -\infty < y < \infty\right\}.$$

Theorem 1.10 *Under the hypothesis of Theorem 1.9,*

a) for any $-\infty < y_1 < y_2 < \infty$,

$$\left(\frac{Z_n(\sqrt{n}\sigma y_1)}{Z_n}, \frac{Z_n(\sqrt{n}\sigma y_2)}{Z_n}\right) \xrightarrow{d} \big(\delta_1(\Phi(y_1)), \delta_2(\Phi(y_2))\big)$$

which takes values $(0,0)$, $(0,1)$ and $(1,1)$ with probabilities $1 - \Phi(y_2)$, $\Phi(y_2) - \Phi(y_1)$ and $\Phi(y_1)$, respectively.

b) for any $-\infty < y_1 < y_2 < \ldots < y_k < \infty$*,*

$$\left(\frac{Z_n(\sqrt{n}\sigma y_i)}{Z_n} : 1 \le i \le k\right) \xrightarrow{d} (\delta_1, \ldots, \delta_k)$$

where each δ_i *is* 0 *or* 1 *and further* $\delta_i = 1 \Rightarrow \delta_j = 1$ *for* $j \ge i$ *and*

$$\begin{aligned} &P(\delta_1 = 0, \delta_2 = 0, \ldots, \delta_{j-1} = 0, \delta_j = 1, \ldots, \delta_k = 1) \\ &= P(\delta_{j-1} = 0, \delta_j = 1) = \Phi(y_j) - \Phi(y_{j-1}). \end{aligned}$$

For proofs of Theorems 1.9 and 1.10, see Athreya and Hong [5].

Theorem 1.11 *If* Y_n *is the position of a randomly chosen individual in the nth generation, then in all cases (as long as* $p_0 = 0$*), given the tree* $\mathcal{T}$*, for all* $y \in \mathbb{R}$*,*

$$P(Y_n \le \sqrt{n}\sigma y \mid \mathcal{T}) \xrightarrow{d} \delta_y \sim \mathit{Bernoulli}(\Phi(y))$$

This is so since

$$P(Y_n \le \sqrt{n}\sigma y \mid \mathcal{T}) = \frac{Z_n(\sqrt{n}\sigma y)}{Z_n}$$

and this in turn implies, for all $y \in \mathbb{R}$,

$$P(Y_n \le \sqrt{n}\sigma y) \to \Phi(y).$$

Remark 1.1 Theorem 1.8 holds under the following weaker assumption about π_k, the distribution of $(X_{k,1}, X_{x,2}, \ldots, X_{k,k})$, that does not require $\{X_{k,1}\}_{k \ge 1}$ to be identically distributed. It suffices to assume:

i) For $k \ge 1$, $(X_{k,1}, X_{x,2}, \ldots, X_{k,k})$ has a distribution that is invariant under permutation.
ii) If $\{p_k\}_{k \ge 1}$ is the offspring distribution with

$$\sum_{k=1}^{\infty} p_k E X_{k,1}^2 < \infty, \quad 1 < m = \sum_{k=1}^{\infty} k p_k < \infty, \quad p_0 = 0.$$

Now let $\mu = \sum\limits_{k=1}^{\infty} p_k E X_{k,1} < \infty$, $\sigma^2 = \sum_{k=1}^{\infty} p_k E X_{k,1}^2 - \mu^2$.

Theorem 1.12 *Let* $\zeta_n \equiv \{x_{n,1}, x_{n,2}, \ldots, x_{n,Z_n}\}$ *be as in Theorem 1.8. Under the above assumptions, the following holds: for* $y \in \mathbb{R}$,

$$\begin{aligned}\frac{Z_n(n\mu + y\sigma\sqrt{n})}{Z_n} &\equiv \frac{1}{Z_n}\sum_{i=1}^{Z_n} I(x_{n,i} \le n\mu + y\sigma\sqrt{n}) \\ &\to \Phi(y) \quad \textit{in mean square.}\end{aligned}$$

1.5.3 Application to Energy Cascades

Consider a particle that under goes fission. Assume each particle spits into a random number of new particles with distribution $\{p_k\}_{k\ge1}$. Assume that the energy x of the parent is split to $\{xY_{k,1}, xY_{k,2}, \ldots, xY_{k,k}\}$ for each of the offspring particle if the parent splits into k offspring particles. Then the energy e_{n,I_n} of a particle I_n in the nth generation can be represented as

$$x_0 Y_{u_1} Y_{u_2} \ldots Y_{u_n}$$

where u_n, u_{n-1},. . ., u_1 are the addresses of the individual I_n and its ancestors and x_0 is the energy of the ancestor 1.

Assume Y_{u_i}'s are independent. Clearly, the distribution of Y_{u_i} depends on the number of offspring of individual u_{i-1} and

$$\left\{ \log e_{n,I_n}, I_n \in n\text{th generation} \right\}$$

is a branching random walk.

So, from Theorem 1.12, one gets the following.

Theorem 1.13 *Let* $\{X_{k,i} \equiv \log Y_{k,i} : 1 \le i \le k\}_{k\ge1}$ *and* $\{p_k\}_{k\ge1}$ *satisfy the conditions of Theorem 1.12. Then, for* $y \in \mathbb{R}$, *as* $n \to \infty$,

$$\begin{aligned}\frac{Z_n(n\mu + y\sigma\sqrt{n})}{Z_n} &\equiv \frac{1}{Z_n}\sum_{i=1}^{Z_n} I(\log e_{n,i} \le n\mu + y\sigma\sqrt{n}) \\ &\to \Phi(y) \quad \textit{in mean square.}\end{aligned}$$

1.6 Scaling Limits of Bellman-Harris Processes with Age Dependent Markov Motion: Supercritical and Critical Cases

Suppose we are given:

i) an offspring distribution $\{p_j\}_{j\geq 1}$ on $\mathbb{N}^+$,
ii) a lifetime distribution $G(\cdot)$ on $(0, \infty)$ and non-lattice,
iii) a real-valued Markov process $\eta(\cdot)$ on $[0, \infty)$ with $\eta(0) = 0$.

First, generate a Bellman-Harris tree $\mathcal{T}$ with offspring distribution $\{p_j\}_{j\geq 0}$ and lifetime distribution $G(\cdot)$ and an initial population at $t = 0$ of size Z_0.

Now, suppose that the initial population is located at $x_{0,i}$, $i = 1, 2, \ldots, Z_0$ and with ages $a_{0,i}$, $i = 1, 2, \ldots, Z_0$.

Assume each individual moves during its lifetime of length L according to Markov process $\{x + \eta(t) : 0 \leq t \leq L\}$.

That is, if an individual is born at time τ and at location x and has lifetime L, then its movement

$$\big\{X(t) : \tau \leq t < \tau + L\big\}$$

is distributed as

$$\big\{x + \eta(t - \tau) : \tau \leq t < \tau + L\big\}$$

where $\{\eta(\cdot)\}$ is a real-valued Markov process on $[0, \infty)$ with $\eta(0) = 0$.

Assume that, for each individual, the lifetime L, the number of offspring ξ and the movement process $\eta(\cdot)$ are independent and the triplets (L, ξ, η) over all the individuals in the tree are i.i.d.

Let Z_t be the population size at time t and

$$C_t \equiv \big\{(a_{t,i}, x_{t,i}) : 1 \leq i \leq Z_t\big\}$$

be the age and position configuration of all the individuals alive at time t.

The object of study is the point process $\{C_t : t \geq 0\}$.

Theorem 1.14 (Supercritical Case (Athreya et al. [8])) *Let* $p_0 = 0$, $1 < m < \infty$. *Let* $E\eta(0) \equiv 0$, $v(t) \equiv E\eta^2(t) < \infty$, $\sup_{0\leq s\leq t} v(s) < \infty$ *and*

$$\psi_\alpha \equiv \int_{[0,\infty)} e^{-\alpha s} v(s) dG(s) < \infty$$

where $0 < \alpha < \infty$ *is the Malthusian parameter defined by*

$$m \int_{[0,\infty)} e^{-\alpha s} dG(s) = 1.$$

Let (a_t, X_t) *be the age and position of a randomly chosen individual at time t. Then*

a)

$$\left(a_t, \frac{X_t}{\sqrt{t}}\right) \xrightarrow{d} (U, V)$$

where U and V are independent and U has pdf proportional to $e^{-\alpha x}\big(1 - G(x)\big)$ *on* $(0,\infty)$ *and V is* $N\left(0, \frac{\psi_\alpha}{\mu_\alpha}\right)$ *where* $\mu_\alpha = m \int_0^\infty x e^{-\alpha x}\big(1 - G(x)\big)dx.$

b) Let

$$Y_y(A \times B) = \frac{1}{Z_t} \sum_{i=1}^{Z_t} I_{A\times B}\big(a_{t,i}, \frac{x_{t,i}}{\sqrt{t}}\big)$$

be the scaled empirical measure of $C_t \equiv \big\{(a_{t,i}, x_{t,i}) : 1 \le i \le Z_t\big\}$.

Then, $Y_t \xrightarrow{d} (U, V)$*, where U and V are as in a).*

The proof of this depends on the following results of independent interest.

Proposition 1.3 *Let* M_t *be the generation number of a randomly chosen individual from* Z_t *(those alive at time t). Let* $\{L_{t,i} : 1 \le t \le M_t\}$ *be the lifetimes of the ancestors of this individual. Then*

a) as $t \to \infty$,

$$\frac{M_t}{t} \to \frac{1}{\mu_\alpha} \quad \text{w.p.1.}$$

b) for any $h : [0,\infty) \to \mathbb{R}$ *Borel measurable and* $\int_{[0,\infty)} |h(x)| e^{-\alpha x} dG(x) < \infty$, $0 < \alpha < \infty$,

$$P\left(\left|\frac{1}{M_t}\sum_{i=1}^{M_t} h(L_{t,i} - c_\alpha(h)\right| > \epsilon\right) \to 0 \quad \text{as } t \to \infty.$$

where $c_\alpha(h) = m \int_{[0,\infty)} h(x) e^{-\alpha x} dG(x).$

Both these results depend on a size-biasing estimate of a large deviation result, namely,

Proposition 1.4 *Let* $\{N(t) : t \geq 0\}$ *be a renewal process generated by G. Let* $1 < m < \infty$ *and* α *be the Malthusian parameter, i.e.,* $m \int_{[0,\infty)} e^{-\alpha x} dG(x) = 1$. *Then, for* $\forall \epsilon > 0$,

$$e^{-\alpha t} E\left(m^{N(t)} I\left(\left| \frac{N(t)}{t} - \frac{1}{\mu_\alpha} \right| > \epsilon \right) \right) = 0$$

where $\mu_\alpha = m \int_0^\infty x e^{-\alpha x} dG(x)$.

Note that since

$$\frac{N(t)}{t} \to \frac{1}{\mu} \quad \text{w.p.1}$$

where $\mu = \int_{[0,\infty)} x dG(x)$, the event

$$\left| \frac{N(t)}{t} - \frac{1}{\mu_\alpha} \right| > \epsilon$$

is an event of large deviation.

Theorem 1.15 (Critical Case) *Let* $m = 1$, $\sum_{j=1}^{\infty} j^2 p_j < \infty$, $E\eta(t) \equiv 0$, $v(t) = E\eta^2(t) < \infty$, $\sup_{0 \leq s \leq t} v(s) < \infty$, $\forall t$, *and*

$$\psi = \int_{[0,\infty)} v(s) dG(s) < \infty.$$

Let $A_t \equiv \{Z_t > 0\}$. *Then, conditioned on* A_t, *the random vector*

$$\left(a_t, \frac{X_t}{t} \right)$$

for a randomly chosen individual converges as $t \to \infty$ *in distribution to* (U, V) *where U and V are independent with U having a pdf* $\frac{1}{\mu}\big(1 - G(\cdot)\big)$ *on* $(0, \infty)$ *and* $V \sim N\big(0, \frac{\psi}{\mu}\big)$.

Theorem 1.16 *Assume the hypothesis of Theorem 1.15. Then, conditioned on* $A_t \equiv \{Z_t > 0\}$*, the empirical measure*

$$Y_t(A \times B) \equiv \frac{1}{Z_t} \sum_{i=1}^{Z_t} I_{A\times B}\big(a_{t,i}, \frac{X_{t,i}}{\sqrt{t}}\big)$$

converges as $t \to \infty$ *in distribution to a random measure* ν *characterized by its moment sequence*

$$m_k(\varphi) \equiv E\big(\langle \nu, \varphi\rangle\big)^k$$

where $\varphi \in C_b^+(\mathbb{R}^+ \times \mathbb{R})$.

The $m_k(\varphi)$ can be expressed in terms of the coalescence times of k randomly chosen individuals alive at time t.

The proof depends on the following results.

Proposition 1.5 *Let* $m = 1$, $\sum_{j=1}^{\infty} j^2 p_j < \infty$, $G(\cdot)$ *non-lattice. Then*

i) $\forall\, \epsilon > 0$

$$P\left(\left|\frac{M_t}{t} - \frac{1}{\mu}\right| > \epsilon \,\middle|\, Z_t > 0\right) \to 0 \quad \text{as } t \to \infty.$$

ii) the coalescence time $\tau_{2,t}$ *of two randomly chosen individuals from time* t *(conditioned on* $Z_t > 0$*) satisfies*

$$\lim_{t\to\infty} P\left(\frac{\tau_{2,t}}{t} \le x \mid Z_t > 0\right) = H(x) \quad \text{exists}$$

for all $0 \le x \le 1$.

iii) A similar result for the convergence of coalescence of k *individuals.*

1.7 Some Extensions and Open Problems

The results in Sect. 1.4 are for the coalescence problem for discrete time single type Galton-Watson processes. Some extension of these have been obtained by J-I Hong. For the discrete time multitype branching processes, the coalescence problem has been addressed in J-I Hong's Ph.D. thesis in the mathematics department, Iowa State University. Some of these results appear in Athreya and Hong [6] and Hong [11]. Some open problems are as follows.

Problem 1 The problem is fully open in the multitype case when the means of offspring in one generation are not finite. In particular, the explosive case results of Athreya (Theorem 1.7) for the single type case need to be formulated and established for the multitype case. First one has to establish an analog of Theorem 1.7 i) and ii).

Problem 2 The problem of extending the results of Sect. 1.4 to the age dependent supercritical and subcritical cases (i.e. Bellman-Harris-Sevastyanov processes) has been addressed in the papers by Hong [10] and [12]. The critical and explosive cases are open.

Problem 3 The problems of extending the results for branching random walks in Athreya and Hong [5] to the multitype discrete time explosive case as well as single type Bellman-Harris explosive case are open.

Problem 4 Athreya and Hong [5] have established the finite dimensional convergence of the normalized position process in the discrete type single type both in the supercritical and explosive cases. The week convergence in the Skorokhod space is open. Only tightness needs to be established.

Problem 5 The corresponding problem (as in Problem 4) for the multitype and Bellman-Harris explosive cases are also open.

Problem 6 Athreya et al. [8, 9], and Hong [12] have established some results for the normalized position process for branching Markov processes for supercritical and critical Bellman-Harris cases using same idea from coalescence. The corresponding problem for the explosive case is fully open.

Some more open problems may be found in Lambert [14] and Legall [15]. Also see the works presented at the workshop by Professors Abrahms and Delmas [1], and Kersting (see http://branching.unex.es/wbpa15/index.htm) [13].

References

1. Abrahms, R., Delmas, J.-F.: Local limits of conditioned Galton-Watson trees: the condensation case. Electron. J. Probab. **19**, Article 56, 1–29 (see also Article 2, 1–19) (2014)
2. Athreya, K.B.: Random walks. In: The Legacy of Alladi Ramakrishnan in the Mathematical Sciences, pp. 337–349. Springer, New York (2010)
3. Athreya, K.B.: Coalescence in the recent past in rapidly growing populations. Stoch. Process. Appl. **122**(11), 3757–3766 (2012)
4. Athreya, K.B.: Coalescence in critical and subcritical Galton-Watson branching processes. J. Appl. Probab. **49**(3), 627–638 (2012)
5. Athreya, K.B., Hong, J.-I.: An application of the coalescence theory to branching random walks. J. Appl. Probab. **50**(3), 893–899 (2013)
6. Athreya, K.B., Hong, J.-I.: Coalescence on critical and subcritical multi-type branching processes. (Submitted)
7. Athreya, K.B., Ney, P.: Branching Processes. Dover, Mineola (2004)

8. Athreya, K.B., Athreya S.B., Iyer, S.-N: Supercritical age dependent branching Markov processes and their scaling limits. Bernoulli **17**, 135–154 (2011)
9. Athreya, K.B., Athreya S.B., Iyer, S.-N.: Critical age dependent branching Markov processes and their scaling limits. In: Math. Proc. I. A. S. (2012)
10. Hong, J.-I.: Coalescence on subcritical Bellman-Harris age-dependent branching processes. J. Appl. Probab. **50**(2), 576–591 (2013)
11. Hong, J.-I.: Coalescence on supercritical multi-type branching processes. Sankhya A **77**(1), 65–78 (2015)
12. Hong, J.-I.: Coalescence on supercritical Bellman-Harris age-dependent branching processes. (Submitted)
13. Kersting, G., Schweinsberg, J., Wakolbinger, A.: The evolving Beta coalescent. Electron. J. Probab. **19**, 1–27 (2014)
14. Lambert, A.: Coalescence time for the branching process. Adv. Appl. Probab. **35**, 1071–1089 (2003)
15. Legall, J.-F.: Ito's extension theory and random trees. Stoch. Process. Appl. **120**, 721–749 (2010)

Chapter 2
A Multi-Type Λ-Coalescent

Robert C. Griffiths

Mathematics Subject Classification (2000): 60J80, 60J75, 92D15

2.1 Introduction

The Λ-coalescent is a random tree back in time which has multiple merger rates for a specific $2 \leq k \leq n$ edges coalescing while n edges in the tree of

$$\lambda_{nk} = \int_0^1 x^k (1-x)^{n-k} \frac{\Lambda(dx)}{x^2},\ k \geq 2, \tag{2.1}$$

where Λ is a non-negative measure on $[0, 1]$. After coalescence there are $n - k + 1$ edges in the tree. The form of the rates (2.1) is necessary for a consistency condition that a subtree of m individuals chosen at random from a larger sample of individuals is distributed as a Λ-coalescent tree with the same rates (2.1). The process is often regarded as having a state space on the set of partitions Π_∞ of the positive integers. The leaves of an infinite-leaf Λ-coalescent tree at time $t = 0$ are labelled with singleton sets $\{1\}, \{2\}, \ldots$ and edges at time t are labelled by sets in $\Pi_\infty(t)$. The number of blocks at time t is the number of sets in the partition $\Pi_\infty(t)$, denoted by $|\Pi_\infty(t)|$, which is the same as the number of edges in the tree at time t. If there are n edges at time t, and k merge at t^+, then a new partition is formed by taking the union of the k partition blocks in the merger for the parent block at t^+. This occurs at rate λ_{nk}. The Λ-coalescent is said to *come down from infinity* if for all $t > 0$,

---

R.C. Griffiths (✉)
University of Oxford, 1 South Parks Road, Oxford, UK
e-mail: griff@stats.ox.ac.uk

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_2

$P(|\Pi_\infty(t)| < \infty) = 1$, which is equivalent to an infinite-leaf Λ-coalescent tree at $t = 0$ having a finite number of edges at any time $t > 0$ back with probability 1. The Λ-coalescent process was introduced in [5, 9, 11] and has been extensively studied [1, 10]. Bertoin and Le Gall [2] showed that the Λ-coalescent comes down from infinity under the same condition that the continuous state branching process becomes extinct in finite time, that is when

$$\int_1^\infty \frac{dq}{\psi(q)} < \infty,$$

where the Laplace exponent

$$\psi(q) = \int_0^1 \left(e^{-qy} - 1 + qy\right)y^{-2}\Lambda(dy).$$

The coalescent process is a moment dual to the Λ-Fleming-Viot process. See for example [6]. There is a distinction between an untyped coalescent process and a typed process such as in [7].

The simplest form of a d-dimensional generator equation of a Λ-Fleming-Viot process $\{\boldsymbol{X}(t)\}_{t\geq 0}$ is

$$\mathcal{L}_\Lambda g(\boldsymbol{x}) = \int_{[0,1]} \sum_{i\in[d]} x_i \left(g(\boldsymbol{x}(1-y) + y\boldsymbol{e}_i) - g(\boldsymbol{x})\right) \frac{\Lambda(dy)}{y^2} \tag{2.2}$$

where $[d] = \{1, 2, \dots, d\}$ and $g \in C^2([0, 1])$. $\boldsymbol{X}(t)$ is a vector of relative frequencies of the d types at time t in a population. If Λ has no atom at zero then the process is a jump process where, reading from the generator equation, an individual of type i gives birth with probability x_i, and has y offspring at rate $\Lambda(dy)/y^2$. The total jump rate $\int_{[0,1]} \Lambda(dy)/y^2$ may be finite or infinite. The frequencies are then rescaled to add to 1. A general reference is [3]. If Λ has an atom of a at zero then $\{\boldsymbol{X}(t)\}_{t\geq 0}$ has a Wright-Fisher diffusive component with part of the generator

$$\frac{a}{2} \sum_{i,j\in[d]} x_i(\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j}.$$

Griffiths [8] shows the following representation for the generator (2.2). Let $\Lambda = F$ be a probability measure, V be a uniform random variable on [0, 1], U a random variable on [0, 1] with density $2u$, $0 < u < 1$ and $W = YU$, where Y has distribution F and V, U, Y are independent. Denote the first and second partial derivatives of a function $g(\boldsymbol{x})$ in $C^2(\Delta)$, $\Delta = \{\boldsymbol{x} : x_i \geq 0, x_1 + \cdots + x_d = 1\}$, by $g_i(\boldsymbol{x})$ and $g_{ij}(\boldsymbol{x})$, $i, j \in [d]$. Then

$$\mathcal{L}_F g(\boldsymbol{x}) = \frac{1}{2} \sum_{i,j=1}^{d} x_i(\delta_{ij} - x_j)\mathbb{E}\Big[g_{ij}\big(\boldsymbol{x}(1 - W) + WV\boldsymbol{e}_i\big)\Big],$$

where expectation $\mathbb{E}$ is taken over V, W. The probabilistic connection with a Wright-Fisher process is not clear. Mutation may be considered to occur at random in the population. If mutation is parent independent with rates θ_i to type i, $i \in [d]$, then a generator equation for the process is (2.2) with an additional term $\sum_{i\in[d]}(\theta_i - |\boldsymbol{\theta}|x_i)g_i(\boldsymbol{x})$ added. If mutation occurs within families on reproduction then a multi-type process arises with different reproduction measures depending on the parents' type. These measures are multi-type, instead of the single type measures in (2.2). There are two different ways to approach adding selection as well, either within the population independently of reproduction, or in families on reproduction.

This paper is a review paper containing a selection of material from [4, 7] explaining multi-type Lambda–Fleming–Viot processes and their dual coalescent processes with mutation and selection without detail of proofs.

2.2 Duality

The simplest duality between the Fleming–Viot process with generator $\mathcal{L}_\Lambda$ and the Λ-coalescent occurs considering the genealogy back in time of a sample of ξ individuals of the same type. Duality in a typed sample is more complex because in a model with no mutation coalescence can only occur between individuals of the same type. The duality is weak duality via a generator equation. To see how this works consider $\mathcal{L}_\Lambda$ acting on x_1^ξ.

$$\begin{aligned}
&\mathcal{L}_\Lambda x_1^\xi \\
&= \int_{[0,1]} \Big(x_1\big(x_1(1-y)+y\big)^\xi + (1-x_1)\big(x_1(1-y)\big)^\xi - x_1^\xi\Big)\,\frac{\Lambda(dy)}{y^2} \\
&= \int_{[0,1]} \left(\sum_{k=2}^{\xi}\binom{\xi}{k} y^k(1-y)^{\xi-k}x_1^{\xi-k+1} - \big(1-(1-y)^\xi - \xi y(1-y)^{\xi-1}\big)x_1^\xi\right)\frac{\Lambda(dy)}{y^2} \\
&= \mathcal{L}^* x_1^\xi,
\end{aligned}$$

where $\mathcal{L}^*$ is a generator acting on functions of $\xi \in \mathbb{Z}_+$ of a process $\{\xi(t)\}_{t\geq 0}$ back in time which has rates

$$\begin{aligned}
q_{\xi,\xi-k+1} &= \int_{[0,1]} \binom{\xi}{k} y^k(1-y)^{\xi-k}\frac{\Lambda(dy)}{y^2},\ k = 2,\dots,\xi \\
q_{\xi,\xi} &= -\int_{[0,1]} \big(1-(1-y)^\xi - \xi y(1-y)^{\xi-1}\big)\frac{\Lambda(dy)}{y^2}.
\end{aligned}$$

These are the same rates as (2.1). The duality equation is

$$\mathbb{E}_{x_1}\big[X_1(t)^{\xi}\big] = \mathbb{E}_{\xi}\big[x_1^{\xi(t)}\big],$$

with $X_1(0) = x_1$, $\xi(0) = \xi$ and where expectation on the left is with respect to $X_1(t)$ and on the right with respect to $\xi(t)$.

2.2.1 Duality with a Stationary Measure

If there is a stationary measure for a process then a particular technique can be used to find a typed dual process. Let $\{\boldsymbol{X}(t)\}_{t\geq 0}$ be a d-dimensional Markov process with generator $\mathcal{L}$ such that $\mathbb{E}\big[\mathcal{L}f\big(\boldsymbol{X}(t)\big)\big] = 0$, under a stationary measure, for all suitable functions f in the domain of the generator. Let $f_{\boldsymbol{\xi}}(\boldsymbol{x})$ be test functions that will be used to construct a dual. For example

$$f_{\boldsymbol{\xi}}(\boldsymbol{x}) = \prod_{j\in[d]} x_j^{\xi_j},\ \boldsymbol{\xi} \in \mathbb{Z}_+^d.$$

Scale by taking

$$g_{\boldsymbol{\xi}}(\boldsymbol{x}) = \frac{f_{\boldsymbol{\xi}}(\boldsymbol{x})}{\mathbb{E}\big[f_{\boldsymbol{\xi}}(\boldsymbol{X})\big]}, \tag{2.3}$$

where expectation is taken in the stationary distribution of $\boldsymbol{X}$. Rearrange the generator equation as

$$\mathcal{L}g_{\boldsymbol{\xi}}(\boldsymbol{x}) = \sum_{\boldsymbol{\chi}} q(\boldsymbol{\xi},\boldsymbol{\chi})\big(g_{\boldsymbol{\chi}}(\boldsymbol{x}) - g_{\boldsymbol{\xi}}(\boldsymbol{x})\big) \tag{2.4}$$

so that $q(\boldsymbol{\xi},\boldsymbol{\chi}) \geq 0$, $\boldsymbol{\chi} \neq \boldsymbol{\xi}$. Set $q(\boldsymbol{\chi},\boldsymbol{\chi}) = -\sum_{\boldsymbol{\chi}\neq\boldsymbol{\xi}} q(\boldsymbol{\xi},\boldsymbol{\chi})$, then Q is a rate matrix of a dual process $\{\boldsymbol{\xi}(t)\}_{t\geq 0}$. The scaling (2.3) is critical in allowing the construction (2.4). The dual equation is

$$\mathbb{E}_{\boldsymbol{x}}\big[g_{\boldsymbol{\xi}}(\boldsymbol{X}(t)\big] = \mathbb{E}_{\boldsymbol{\xi}}\big[g_{\boldsymbol{\xi}(t)}(\boldsymbol{x})\big]$$

with expectation taken on the left forward in time with respect to $\boldsymbol{X}(t)$, with $\boldsymbol{X}(0) = \boldsymbol{x}$ and on the right backward in time with respect to $\boldsymbol{\xi}(t)$, with $\boldsymbol{\xi}(0) = \boldsymbol{\xi}$.

For example consider the generator (2.2) with the addition of parent independent mutation occurring independently from reproduction in the population. The generator is then specified by

$$\mathcal{L}g(\boldsymbol{x}) = \int_{[0,1]} \sum_{i\in[d]} x_i \left(g(\boldsymbol{x}(1-y)+y\boldsymbol{e}_i) - g(\boldsymbol{x})\right) \frac{\Lambda(dy)}{y^2}$$
$$+ \sum_{i\in[d]} (\theta_i - \theta x_i) \frac{\partial}{\partial x_i} g(\boldsymbol{x}),$$

where $\theta = \sum_{i\in[d]} \theta_i$. A stationary measure exists although its form is unknown. Denote $\boldsymbol{x}^{\boldsymbol{\xi}} = \prod_{i\in[d]} x_i^{\xi_i}$. The sampling distribution in a stationary population will be denoted by

$$\mathcal{M}(\boldsymbol{\xi}) = \mathbb{E}\left[\binom{|\boldsymbol{\xi}|}{\boldsymbol{\xi}} \boldsymbol{X}^{\boldsymbol{\xi}}\right].$$

Test functions for the dual process are

$$g_{\boldsymbol{\xi}}(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\boldsymbol{\xi}}}{\mathbb{E}[\boldsymbol{X}^{\boldsymbol{\xi}}]} = \frac{\binom{|\boldsymbol{\xi}|}{\boldsymbol{\xi}}\boldsymbol{x}^{\boldsymbol{\xi}}}{\mathcal{M}(\boldsymbol{\xi})}.$$

Then

$$\begin{aligned}
\mathcal{L}\boldsymbol{x}^{\boldsymbol{\xi}} &= \int_{[0,1]} \sum_{i\in[d]} x_i \left((x_i(1-y)+y)^{\xi_i}(1-y)^{|\boldsymbol{\xi}|-\xi_i} \boldsymbol{x}^{\boldsymbol{\xi}-\boldsymbol{e}_i\xi_i} - \boldsymbol{x}^{\boldsymbol{\xi}} \right) \frac{F(dy)}{y^2} \\
&\qquad + \sum_{i\in[d]} (\theta_i - \theta x_i)\xi_i \boldsymbol{x}^{\boldsymbol{\xi}-\boldsymbol{e}_i} \\
&= \int_{[0,1]} \sum_{i\in[d]} \sum_{l\geq 2} \binom{\xi_i}{l} y^l (1-y)^{|\boldsymbol{\xi}|-l} \frac{F(dy)}{y^2} \boldsymbol{x}^{\boldsymbol{\xi}-\boldsymbol{e}_i(l-1)} + \sum_{i\in[d]} \theta_i \xi_i \boldsymbol{x}^{\boldsymbol{\xi}-\boldsymbol{e}_i} \\
&\qquad - \left[\int_{[0,1]} \sum_{i\in[d]} \left(1-(1-y)^{|\boldsymbol{\xi}|} - |\boldsymbol{\xi}|y(1-y)^{|\boldsymbol{\xi}|-1}\right) \frac{F(dy)}{y^2} + \theta|\boldsymbol{\xi}| \right] \boldsymbol{x}^{\boldsymbol{\xi}}
\end{aligned} \tag{2.5}$$

The dual system $\{\boldsymbol{\xi}(t)\}_{t\geq 0}$ is read off from (2.5) as a multi-type death process with rates for $l \geq 2$ of

$$q(\boldsymbol{\xi}, \boldsymbol{\xi} - \boldsymbol{e}_i(l-1)) = \frac{\xi_i + 1 - l}{|\boldsymbol{\xi}| + 1 - l} \int_{[0,1]} \binom{|\boldsymbol{\xi}|}{l} y^l (1-y)^{|\boldsymbol{\xi}|-l} \frac{F(dy)}{y^2} \times \frac{\mathcal{M}(\boldsymbol{\xi} - \boldsymbol{e}_i(l-1))}{\mathcal{M}(\boldsymbol{\xi})},$$

$$q(\boldsymbol{\xi}, \boldsymbol{\xi} - \boldsymbol{e}_i) = \theta_i |\boldsymbol{\xi}| \frac{\mathcal{M}(\boldsymbol{\xi} - \boldsymbol{e}_i)}{\mathcal{M}(\boldsymbol{\xi})}. \tag{2.6}$$

Mutations occur at random on coalescent lines in the dual process.

Models with selection are more involved and lead to coalescing branching dual processes back in time.

2.3 Multi-type Moran Model with Viability Selection

Sample path duality is easier to see in a Moran model where there is an explicit construction. There is a fixed population of N individuals with a type space of $[d]$. Reproduction events occur at rate λ. On reproduction a juvenile offspring are produced with probability distribution $\{r_a\}_{a\in[N-1]}$.

The distribution of the numbers of types $\boldsymbol{B}$ which survive to maturity from a juvenile offspring of a type i parent is $\{Q_{ia}(\boldsymbol{B})\}$. The probability that a type i parent produces $\boldsymbol{B}$ mature offspring is then

$$p_i(\boldsymbol{B}) = \sum_{a=|\boldsymbol{B}|}^{N-1} r_a Q_{ia}(\boldsymbol{B}). \tag{2.7}$$

$|\boldsymbol{B}|$ individuals are chosen to be killed, excluding the parent. This is viability selection and mutation where offspring are thinned according to $Q_{ia}(\cdot)$.

Two possible biological models which incorporate mutation and selection are the following:

1. Mutation occurs independently to juvenile offspring according to a $d \times d$ type transition matrix U. Mutation is followed by viability selection to form the mature offspring $\boldsymbol{B}$.
2. Selection occurs at the parental level, reducing the a juvenile offspring to $b \leq a$. Mutation then occurs according to a type transition matrix U to form the mature offspring $\boldsymbol{B}$, with $|\boldsymbol{B}| = b$.

An emphasis is given to the first model in this paper. The Moran model does not know the internal mechanism of reproduction. It only sees multi-type reproduction described by $p_i(\boldsymbol{B})$, so while viability selection is used as a mechanism here, other mechanisms, such as fecundity selection, may give the same probabilities $p_i(\boldsymbol{B})$. The hypergeometric distribution of a sample $\boldsymbol{\gamma}$, with $|\boldsymbol{\gamma}|$ fixed, taken at random without replacement from a population configuration z, will be denoted by

$$\mathcal{H}(\boldsymbol{\gamma} \mid z) = \frac{\prod_{j\in[d]} \binom{z_j}{\gamma_j}}{\binom{|z|}{|\boldsymbol{\gamma}|}}.$$

The generator of the process $\{\boldsymbol{Z}(t)\}_{t\geq 0}$ acting on a function $f(z), z \in \mathbb{Z}_+^d$, $|z| = N$ is

$$\mathcal{L}f(z) = \lambda \sum_{\mathcal{R}(i,\boldsymbol{B},\boldsymbol{\gamma})} \frac{z_i}{N} p_i(\boldsymbol{B})\mathcal{H}(\boldsymbol{\gamma} \mid z - \boldsymbol{\gamma})\big(f(z + \boldsymbol{B} - \boldsymbol{\gamma}) - f(z)\big),$$

with summation over the region

$$\mathcal{R}(i, \boldsymbol{B}, \boldsymbol{\gamma}) = \big\{i \in [d], \boldsymbol{B} \in \Delta_{N-1}, \boldsymbol{\gamma} \in \Delta_{N-1}, |\boldsymbol{\gamma}| = |\boldsymbol{B}|\big\},$$

where the notation $\Delta_k = \{\boldsymbol{\xi} \in \mathbb{Z}_+^d, |\boldsymbol{\xi}| \leq k\}, k \in \mathbb{Z}_+$, is used.

2.3.1 Dual Process in the Moran Model

Real and virtual births have to be taken into consideration to obtain a proper dual process back in time. In a reproduction event:

$$a \text{ juvenile offspring} \longrightarrow \begin{cases} \boldsymbol{B} \text{ mature offspring: } \textit{real births} \\ a - |\boldsymbol{B}| \text{ juveniles killed: } \textit{virtual births.} \end{cases}$$

There is a rule for adding virtual lineages corresponding to branching in the genealogy that a parental line should be added if the parent has no real offspring, but at least one virtual offspring placed in ancestral lines. An example is shown in Fig. 2.1.

There are two types in the diagram represented by ▲ and $*$. Time runs down the diagram. The dual process genealogy of a sample of 5 ▲ and 2 $*$ individuals is followed back in time. Black lines of ▲ type and dashed black lines of $*$ type belong to the genealogy of the sample. Gray lines are not part of the genealogy of the sample. Black arrows represent births of ▲ type, and dashed black arrows represent births of $*$ type. Virtual births are represented by dotted arrows. The first event back in time is a birth event from a ▲ parent which has a $*$ birth not affecting the sample genealogy, and a virtual birth which attempts to replace an individual in the sample genealogy. The parent lineage then must be added to the sample genealogy. The next event is a birth event from a $*$ parent which has two births of type $*$,

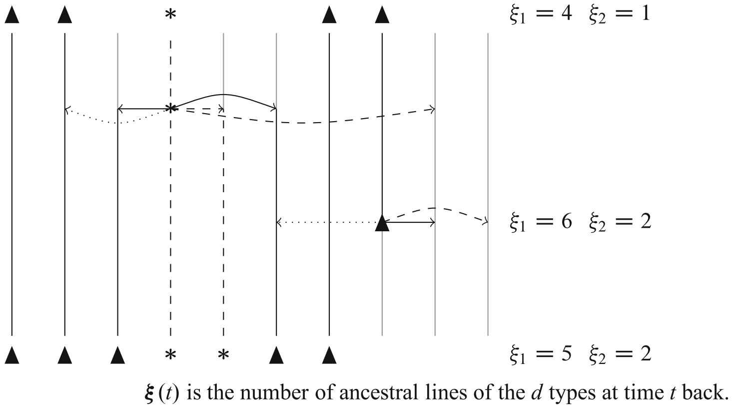


Fig. 2.1 A dual Λ-branching-coalescing graph back in time

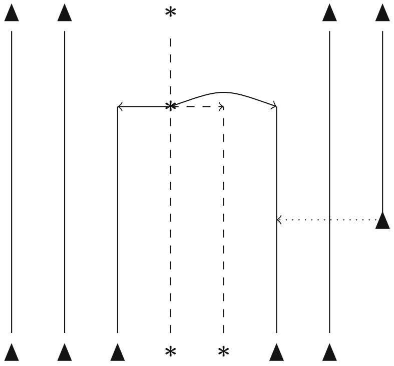


Fig. 2.2 Branching-coalescing graph

two births of type ▲ and a virtual birth. Following which lines are replaced there is a coalescence of three $*$ lines and two ▲ lines. Finally there are four type ▲ and one type $*$ ancestors of the sample. A branching-coalescing graph back in time is extracted from the ancestral lines, shown in Fig. 2.2.

Let $\mathbf{Z}(t)$ be the number of individuals of the d types at time $t \geq 0$ in the Moran model with viability selection. The dual process counting lineages of different types is $\{\boldsymbol{\xi}(t)\}_{t\geq 0}$. Assume a stationary distribution exists for $\mathbf{Z}(t)$. This occurs if there is recurrent mutation between types. Let $\mathcal{H}(\boldsymbol{\xi})$ be the stationary sampling distribution

$$\mathcal{H}(\boldsymbol{\xi}) = \mathbb{E}\left[\frac{\binom{Z_1}{\xi_1}\cdots\binom{Z_d}{\xi_d}}{\binom{N}{|\boldsymbol{\xi}|}}\right]$$

where expectation is over $\mathbf{Z}$.

Theorem 2.1 (Birkner et al. [4]) *$\{\boldsymbol{Z}(t)\}_{t\geq 0}$ is dual to the process $\{\boldsymbol{\xi}(t)\}_{t\geq 0}$ in Δ_N with non-zero multi-type coalescence rates $q(\boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{e}_i - \boldsymbol{l})$ and branching rates $q(\boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{e}_i)$.*

The transition $\boldsymbol{\xi} \to \boldsymbol{\xi} + \boldsymbol{e}_i - \boldsymbol{l}$ occurs when forward in time a type i parent has $\boldsymbol{l}$ real offspring in the ancestral lines. The parent can be inside or outside the ancestral lines. The transition $\boldsymbol{\xi} \to \boldsymbol{\xi} + \boldsymbol{e}_i$ occurs when at least one virtual, but no real offspring are in the ancestral lineages. The transition rates are combinatorial but simplify in a limit model. The rates in Theorem 2.1 are

$$q(\boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{e}_i - \boldsymbol{l}) = \lambda \left[\sum_{\boldsymbol{B} \in \Delta_{N-1}} p_i(\boldsymbol{B}) \mathcal{H}(\boldsymbol{l} \mid \boldsymbol{B} + \boldsymbol{e}_i) \frac{\binom{|\boldsymbol{\xi}|}{|\boldsymbol{l}|}\binom{N-|\boldsymbol{\xi}|}{|\boldsymbol{B}|+1-|\boldsymbol{l}|}}{\binom{N}{|\boldsymbol{B}|+1}} \right] \times \frac{\mathcal{H}(i, \boldsymbol{\xi} - \boldsymbol{l})}{\mathcal{H}(\boldsymbol{\xi})},$$

$$q(\boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{e}_i) = \frac{\lambda}{N}(N - |\boldsymbol{\xi}|) p^*_{i,|\boldsymbol{\xi}|} \times \frac{\mathcal{H}(i, \boldsymbol{\xi})}{\mathcal{H}(\boldsymbol{\xi})},$$

where $\boldsymbol{0} < \boldsymbol{l} \leq \boldsymbol{\xi}, \boldsymbol{l} \neq \boldsymbol{e}_i$, $|\boldsymbol{l}| \geq 1$.

$$\mathcal{H}(i, \boldsymbol{\xi}) = \frac{\xi_i + 1}{|\boldsymbol{\xi}| + 1} \mathcal{H}(\boldsymbol{\xi} + \boldsymbol{e}_i)$$

denotes the probability that the parent chosen in a reproduction event is of type i and a sample of size $|\boldsymbol{\xi}|$ from the remaining individuals in the population has configuration $\boldsymbol{\xi}$. The dual equation is based on falling factorial moments. Denote $z_{[n]} = z(z-1)\cdots(z-n+1)$ and take

$$g_{\boldsymbol{\xi}}(\boldsymbol{z}) = \frac{\prod_{j=1}^{d} z_{i[\xi_i]}}{\mathbb{E}\big[\prod_{j=1}^{d} Z_{i[\xi_i]}\big]}.$$

The duality equation is

$$\mathbb{E}_{\boldsymbol{Z}(0)}\Big[g_{\boldsymbol{\xi}(0)}\big(\boldsymbol{Z}(t)\big)\Big] = \mathbb{E}_{\boldsymbol{\xi}(0)}\Big[g_{\boldsymbol{\xi}(t)}\big(\boldsymbol{Z}(0)\big)\Big],$$

where the expectation on the left is with respect to the distribution of $\boldsymbol{Z}(t)$ and on the right is with respect to the distribution of $\boldsymbol{\xi}(t)$. The probability that a type i parent has no real, but at least one virtual offspring in $\boldsymbol{\xi}$ ancestral lines is $p^*_{i|\boldsymbol{\xi}|}$. These probabilities depend on the details of viability selection. Let

$$p^*_{i|\boldsymbol{\xi}||\boldsymbol{B}|} = \sum_{k \in [N-1]} r_k v_{ik|\boldsymbol{B}|} \left\{ \frac{\binom{N-1-|\boldsymbol{\xi}|}{|\boldsymbol{B}|}}{\binom{N-1}{|\boldsymbol{B}|}} - \frac{\binom{N-1-|\boldsymbol{\xi}|}{k}}{\binom{N-1}{k}} \right\};$$

the probability that there are $|\boldsymbol{B}|$ offspring, of which no real offspring replace individuals in ancestral lineages; then

$$p^*_{i|\boldsymbol{\xi}|} = \sum_{|\boldsymbol{B}|=0}^{N-1} p^*_{i|\boldsymbol{\xi}||\boldsymbol{B}|},$$

where

$$v_{ik|\boldsymbol{B}|} = \sum_{\{\boldsymbol{B}\in\Delta_{N-1},|\boldsymbol{B}| \text{ fixed}\}} Q_{ik}(\boldsymbol{B})$$

is the probability, irrespective of offspring types, that k juvenile offspring from a type i parent survive to $|\boldsymbol{B}|$ mature offspring. Then

$$p_i(|\boldsymbol{B}|) = \sum_{k\in[N-1]} r_k v_{ik|\boldsymbol{B}|}.$$

An identity is

$$p^*_{i|\boldsymbol{\xi}|} = \sum_{|\boldsymbol{B}|\in[N-1]} \big(r_{|\boldsymbol{B}|} - p_i(|\boldsymbol{B}|)\big)\left(1 - \frac{\binom{N-1-|\boldsymbol{\xi}|}{|\boldsymbol{B}|}}{\binom{N-1}{|\boldsymbol{B}|}}\right).$$

Rates in the dual process contain *Bayes' factors*

$$\boldsymbol{\xi} \to \boldsymbol{\xi} - \boldsymbol{l} + \boldsymbol{e}_i \;;\; \mathcal{H}(i, \boldsymbol{\xi} - \boldsymbol{l})/\mathcal{H}(\boldsymbol{\xi})$$
$$\boldsymbol{\xi} \to \boldsymbol{\xi} + \boldsymbol{e}_i \;;\; \mathcal{H}(i, \boldsymbol{\xi})/\mathcal{H}(\boldsymbol{\xi})$$

because of typed lines and time progressing backward. The duality in Theorem 2.1 is a weak duality because it is argued by rearranging the generator acting on factorial moments, however if virtual births are included in a forward model then the duality is a strong duality in that given a path of $\{\boldsymbol{\xi}(t)\}$ the path of $\{\boldsymbol{Z}(t)\}$ can be reconstructed as a line diagram up to permutations of the lines. The rates in Theorem 2.1 can then be argued in a probabilistic way. Virtual births must be included in the forward sample paths so there is a unique dual sample path back in time for each forward sample path. The rate $q(\boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{e}_i - \boldsymbol{l})$ is λ times the probability that given a sample configuration $\boldsymbol{\xi}$, that it arose through a reproduction → mutation → selection event that left a configuration of $\boldsymbol{B}$ mature offspring and a total configuration of $\boldsymbol{l}$ from the parent and $\boldsymbol{B}$ mature offspring are in the sample. Multiple mergers occur because a parent produces multiple offspring types. The rate $q(\boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{e}_i)$ reflecting selection is λ times the conditional probability, given a sample with configuration $\boldsymbol{\xi}$, that a type i parent was chosen, the parent lies outside the sample, and at least one non-viable arrow head but no viable arrow head fell within the sample. The parental line is then added as a virtual lineage.

2.4 Multi-type Λ–Fleming–Viot Process with Mutation and Selection

An analogy to the multi-type Moran model is a Λ–Fleming–Viot process where an individual of type i is chosen to reproduce with probability x_i and has a proportion of multi-type offspring $\boldsymbol{y}$ with rate $G_i(d\boldsymbol{y})/|\boldsymbol{y}|^2$, where G_i are sub-stochastic measures. The model includes mutation and selection in the measures G_i. The generator is described for $g \in C^2(\Delta_\circ^d)$ by

$$\mathcal{L}_{mt} g(\boldsymbol{x}) = \int_{\Delta^d} \sum_{i \in [d]} x_i \left(g(\boldsymbol{x}(1 - |\boldsymbol{y}|) + \boldsymbol{y}) - g(\boldsymbol{x}) \right) \frac{G_i(d\boldsymbol{y})}{|\boldsymbol{y}|^2}, \tag{2.8}$$

where $\Delta^d = \{\boldsymbol{y} \in \mathbb{R}_+^d; |\boldsymbol{y}| \le 1\}$, $\Delta_\circ^d = \{\boldsymbol{x} \in \mathbb{R}_+^d; |\boldsymbol{x}| = 1\}$. $\int_{\Delta^d} G_i(d\boldsymbol{y})/|\boldsymbol{y}| < \infty$ is required for the generator to be well defined. A stationary distribution exists if the multi-type offspring distribution is recurrent. An analogy to a viability structure in the Moran model is to assume that there exists a probability measure F on $[0, 1]$ and sub-stochastic viability measures $V_i(\boldsymbol{w}, \cdot)$ supported on $\{\boldsymbol{y} : \boldsymbol{y} \in \Delta^d; |\boldsymbol{y}| \le |\boldsymbol{w}|\}$ such that

$$\frac{G_i(d\boldsymbol{y})}{|\boldsymbol{y}|^2} = \int_{|\boldsymbol{y}| < |\boldsymbol{w}| < 1} V_i(\boldsymbol{w}, d\boldsymbol{y}) \frac{F(d\boldsymbol{w})}{|\boldsymbol{w}|^2}. \tag{2.9}$$

It is also assumed that

$$K_i(d\boldsymbol{y}) = \frac{F(d\boldsymbol{y}) - G_i(d\boldsymbol{y})}{|\boldsymbol{y}|} \tag{2.10}$$

defines a finite signed measure on Δ^d. The interpretation of (2.9) is that multi-type juvenile offspring from a type i individual are produced at rate $F(d\boldsymbol{w})/|\boldsymbol{w}|^2$, not depending on the type of the parent, then thinned by viability selection according to $V_i(\boldsymbol{w}, d\boldsymbol{y})$. The model with generator (2.8) is valid as it stands, and also can be obtained as a limit from the Moran model where it is assumed that

$$\lambda^{(N)} \sum_{\boldsymbol{B} \in \Delta_{N-1}} p_i(\boldsymbol{B}) f(\boldsymbol{B}/N) \to \int_{\Delta^d} f(\boldsymbol{y}) \frac{G_i(d\boldsymbol{y})}{|\boldsymbol{y}|^2}$$

where $f \in C^2(\Delta^d)$ is such that $|f(\boldsymbol{y})|/|\boldsymbol{y}|^2$ is bounded as $|\boldsymbol{y}| \to 0$. F and $V_i(\boldsymbol{x}, \cdot)$ are analogous to r_a and $Q_{ia}(\cdot)$ in (2.7). In a Wright-Fisher diffusion process the strength of selection is controlled by the selection parameters $\{\sigma_i\}$ and the frequencies change according to a term

$$\sum_{j \in [d]} x_j (\sigma_j - \sum_{k \in [d]} \sigma_k x_k) \frac{\partial}{\partial x_j} \tag{2.11}$$

in the generator. The measures $\{K_i\}$ are analogous to the parameters $\{\sigma_i\}$ however their effect is *much* stronger on the sample paths than occurs with the deterministic change governed by (2.11).

In [7] a Λ–Fleming–Viot process with viability selection has a generator equation

$$\mathcal{L}_\Lambda g(\boldsymbol{x}) = \int_{[0,1]} \sum_{i\in[d]} x_i \left(g(\boldsymbol{x}(1-y) + y\boldsymbol{e}_i) - g(\boldsymbol{x})\right) \frac{G_{1i}(dy)}{y^2},$$

where G_{1i} are one-dimensional sub-stochastic measures, and $g \in C^2([0,1])$. This is a particular case of (2.8) where the measures G_i have mass concentrated on y_i, with $G_{1i}(dy) = G_i(\boldsymbol{e}_i dy)$. Parents of type i give rise to type i offspring in this model. There is a similar definition for the one-dimensional measures K_i as in (2.10). If $K_i(\cdot) = -\sigma_i\delta_\epsilon(\cdot)$ then as $\epsilon \to 0$ the corresponding sequence of generators converges to

$$\begin{aligned}\mathcal{L}_\Lambda g(\boldsymbol{x}) &= \int_{[0,1]} \sum_{i\in[d]} x_i \left(g(\boldsymbol{x}(1-y) + ye_i) - g(\boldsymbol{x})\right) \frac{F(dy)}{y^2} \\ &\quad + \sum_{j\in[d]} x_j(\sigma_j - \sum_{k\in[d]} \sigma_k x_k) \frac{\partial}{\partial x_j} g(\boldsymbol{x}).\end{aligned}$$

If there is no selection in the models (2.8) or (2.9) then $G_i = F$ and K_i are null measures, $i \in [d]$.

Example 2.1 If there is no selection and mutation is parent independent with a mutation rate of ϵ_j, $\sum_{i\neq j} \epsilon_i \leq 1, j \in [d]$ the generator equation (2.8) becomes

$$\begin{aligned}\mathcal{L}_{mt} g(\boldsymbol{x}) = \int_{[0,1]} \sum_{i\in[d]} x_i \Bigg(& g\big(\boldsymbol{x}(1-|\boldsymbol{y}|) + |y|(1-|\boldsymbol{\epsilon}| + \epsilon_i)\boldsymbol{e}_i + |\boldsymbol{y}| \sum_{j\neq i} \epsilon_j \boldsymbol{e}_j\big) \\ & - g(\boldsymbol{x})\Bigg) \frac{F(d|\boldsymbol{y}|)}{|\boldsymbol{y}|^2}.\end{aligned} \tag{2.12}$$

If $|\boldsymbol{\epsilon}| = 1$ the outcome of a jump does not depend on the type of the reproducing individual. Then the stationary distribution is the constant vector $\boldsymbol{\epsilon}$.

Example 2.2 If $d = 2$ in (2.12) then the generator for the frequency of the first type is specified by

$$\begin{aligned}\mathcal{L}_{mt} g(x) = \int_{[0,1]} \Big[& x\big(g\big(x(1-y) + (1-\epsilon_2)y\big) - g(x)\big) \\ & + (1-x)\big(g\big(x(1-y) + \epsilon_1 y\big) - g(x)\big)\Big] \frac{F(dy)}{y^2}.\end{aligned}$$

Example 2.3 A two-type model with Beta measures and no mutation. The generator for X_1 acting on $g(x)$ is

$$x \int_0^1 [g((1-y)x+y) - g(x)] \frac{B_{2-\gamma_1,\gamma_1}(dy)}{y^2} + (1-x) \int_0^1 [g((1-y)x) - g(x)] \frac{B_{2-\gamma_2,\gamma_2}(dy)}{y^2}$$

This model arises naturally from a continuous time Schweinsberg differential rates fecundity birth model or from a model with viability selection for the two types. The parameters $\gamma_1, \gamma_2 \in (0,1)$. The measures $G_i(dy) = B_{2-\gamma_i,\gamma_i}(dy)$, $i = 1,2$. Take $F = G_1$, so then there is no thinning of type 1 individuals. A classical thinning identity is to take $Y_i \sim G_i$, $i = 1,2$ and $U \sim B_{2-\gamma_2,\gamma_2}$ independent of Y_1 such that for $\gamma_2 > \gamma_1$, $Y_2 =^{\mathcal{D}} Y_1 U$. Let $f(y_2 \mid y_1)$ be the density of $Y_2 \mid Y_1 = y_1$, then

$$B_{2-\gamma_2,\gamma_2}(dy_2) = \int_{y_2}^1 f(y_2 \mid y_1) dy_2 B_{2-\gamma_1,\gamma_1}(dy_1)$$

This example also shows that the same process can be obtained by either fecundity selection, or viability selection.

There is a similar dual process to that of Theorem 2.1 in the Moran model, with a simpler form for the rates. The dual process $\{\boldsymbol{\xi}(t)\}_{t\geq 0}$ is again a branching coalescing graph back in time following the genealogy of a sample of individuals back in time. It is assumed that there is a stationary distribution for $\{\boldsymbol{X}(t)\}_{t\geq 0}$ with sampling distribution $\mathcal{M}(\boldsymbol{\xi})$. The probability that a parent chosen in a reproduction event is of type i, and a sample configuration of an additional $|\boldsymbol{\eta}|$ individuals has configuration $\boldsymbol{\eta}$ is

$$\mathcal{M}(i,\boldsymbol{\eta}) = \frac{\eta_i + 1}{|\boldsymbol{\eta}| + 1} \mathcal{M}(\boldsymbol{\eta} + \boldsymbol{e}_i),\ \boldsymbol{\eta} \in \mathbb{Z}_+^d.$$

Theorem 2.2 ([4]) *The multi-type Fleming–Viot process with generator (2.8) is dual to the system of branching and coalescing lineages $\{\boldsymbol{\xi}(t)\}_{t\geq 0}$ which takes values in $\mathbb{Z}_+^d$ and for which the transition rates are, for $\boldsymbol{0} \leq \boldsymbol{l} \leq \boldsymbol{\xi}$, $\boldsymbol{l} \neq \boldsymbol{e}_i$, $|\boldsymbol{l}| \geq 1$.*

$$q(\boldsymbol{\xi}, \boldsymbol{\xi} - \boldsymbol{l} + \boldsymbol{e}_i) = \int_{\Delta^d} \left[\binom{|\boldsymbol{\xi}|}{|\boldsymbol{l}|} \binom{|\boldsymbol{l}|}{\boldsymbol{l}} \prod_{j\in[d]} y_j^{l_j} \times (1 - |\boldsymbol{y}|)^{|\boldsymbol{\xi}|-|\boldsymbol{l}|} \right] \frac{G_i(d\boldsymbol{y})}{|\boldsymbol{y}|^2} \times \frac{\mathcal{M}(i, \boldsymbol{\xi} - \boldsymbol{l})}{\mathcal{M}(\boldsymbol{\xi})},$$

$$q(\boldsymbol{\xi}, \boldsymbol{\xi} + \boldsymbol{e}_i) = \int_{[0,1]} \left(1 - (1 - |\boldsymbol{y}|)^{|\boldsymbol{\xi}|}\right) \frac{K_i(d|\boldsymbol{y}|)}{|\boldsymbol{y}|}$$

$$\times \frac{\mathcal{M}(i,\boldsymbol{\xi})}{\mathcal{M}(\boldsymbol{\xi})},$$

$$q(\boldsymbol{\xi},\boldsymbol{\xi}) = -\Bigg[\int_{\Delta^d}\Big(1-(1-|\boldsymbol{y}|)^{|\boldsymbol{\xi}|}\Big)\frac{F(d|\boldsymbol{y}|)}{|\boldsymbol{y}|^2} - \sum_{i\in[d]}\xi_i y_i(1-|\boldsymbol{y}|)^{|\boldsymbol{\xi}|-1}\frac{G_i(d\boldsymbol{y})}{|\boldsymbol{y}|^2}\Bigg]$$

The duality equation is

$$\mathbb{E}_{\boldsymbol{X}(0)}\left[\frac{\binom{|\boldsymbol{\xi}(0)|}{\boldsymbol{\xi}(0)}\boldsymbol{X}(t)^{\boldsymbol{\xi}(0)}}{\mathcal{M}\big(\boldsymbol{\xi}(0)\big)}\right] = \mathbb{E}_{\boldsymbol{\xi}(0)}\left[\frac{\binom{|\boldsymbol{\xi}(t)|}{\boldsymbol{\xi}(t)}\boldsymbol{X}(0)^{\boldsymbol{\xi}(t)}}{\mathcal{M}\big(\boldsymbol{\xi}(t)\big)}\right],$$

where the expectation on the left hand side is with respect to the distribution of the forwards in time Fleming–Viot process $\{\boldsymbol{X}(t)\}_{t\geq 0}$ *and that on the right is with respect to the backwards in time dual process* $\{\boldsymbol{\xi}(t)\}_{t\geq 0}$.

Example 2.4 In the model of Example 2.2 the dual rates (2.6) for $\boldsymbol{0} \leq \boldsymbol{l} \leq \boldsymbol{\xi}, \boldsymbol{l} \neq \boldsymbol{e}_i$, $i = 1, 2$, $|\boldsymbol{l}| \geq 1$ become

$$q(\boldsymbol{\xi},\boldsymbol{\xi}-\boldsymbol{l}+\boldsymbol{e}_1) = \binom{|\boldsymbol{l}|}{\boldsymbol{l}}\binom{|\boldsymbol{\xi}|}{|\boldsymbol{l}|}(1-\epsilon_2)^{l_1}\epsilon_2^{l_2}\int_{[0,1]} y^{|\boldsymbol{l}|}(1-y)^{|\boldsymbol{\xi}|-|\boldsymbol{l}|}\frac{F(dy)}{y^2} \times \frac{\mathcal{M}(1,\boldsymbol{\xi}-\boldsymbol{l})}{\mathcal{M}(\boldsymbol{\xi})}$$

$$q(\boldsymbol{\xi},\boldsymbol{\xi}-\boldsymbol{l}+\boldsymbol{e}_2) = \binom{|\boldsymbol{l}|}{\boldsymbol{l}}\binom{|\boldsymbol{\xi}|}{|\boldsymbol{l}|}\epsilon_1^{l_1}(1-\epsilon_1)^{l_2}\int_{[0,1]} y^{|\boldsymbol{l}|}(1-y)^{|\boldsymbol{\xi}|-|\boldsymbol{l}|}\frac{F(dy)}{y^2} \times \frac{\mathcal{M}(2,\boldsymbol{\xi}-\boldsymbol{l})}{\mathcal{M}(\boldsymbol{\xi})}.$$

Mutation occurs in families in this example, in contrast to mutation occurring at random along lines in the dual process with rates (2.6).

2.5 Conclusion

This paper discusses multi-type Λ-coalescent processes which arise naturally from Λ-Fleming-Viot processes as dual processes back in time. A general problem is to see if there is an analogue of these population-coalescent pairs of processes in branching processes theory.

References

1. Berestycki, N.: Recent progress in coalescent theory. Ensaios Matematicos **16**, 1–193 (2009)
2. Bertoin, J., Le Gall, J.-F.: Stochastic flows associated to coalescent processes III: limit theorems. Ill. J. Math. **50**, 147–181 (2006)
3. Birkner, B., Blath, J.: Measure-valued diffusions, general coalescents and population genetic inference. In: Trends in Stochastic Analysis, Chap. 12. Cambridge University Press, Cambridge (2009)
4. Birkner, M., Etheridge, A., Griffiths, R., Miller, L.: A multi-type Λ-coalescent with mutation and selection. In preparation (2016)
5. Donnelly, P.J., Kurtz, T.G.: Particle representations for measure-valued population models. Ann. Probab. **24**, 166–205 (1999)
6. Etheridge, A.: Some mathematical models from population genetics. In: École d'Été de Probabilitités de Saint-Flour XXXIX-2009. Springer, Heidelberg (2012)
7. Etheridge, A.M., Griffiths, R.C., Taylor, J.E.: A coalescent dual process in a moran model with genic selection, and the lambda coalescent limit. Theor. Popul. Biol. **78**, 77–92 (2010)
8. Griffiths R.C.: The Lambda-Fleming-Viot process and a connection with Wright-Fisher diffusion. Adv. Appl. Probab. **46**, 1009–1035 (2014)
9. Pitman, J.: Coalescents with multiple collisions. Ann. Probab. **27**, 1870–1902 (1999)
10. Pitman, J.: Combinatorial stochastic processes. In: École d'Été de Probabilitités de Saint-Flour XXXII-2002. Springer, Heidelberg (2006)
11. Sagitov, S.: The general coalescent with asynchronous mergers of ancestral lines. J. Appl. Probab. **36**, 1116–1125 (1999)

Part II
Branching Random Walks

Chapter 3
On the Number of Positive Eigenvalues of the Evolutionary Operator of Branching Random Walk

Ekaterina Antonenko and Elena Yarovaya

Mathematics Subject Classification (2000): 60J80, 60J25, 34G10, 47B39

3.1 Introduction

Let $\mathbb{Z}^d$, called in what follows the d-dimensional integer lattice, be a set of vectors $(n_1, n_2, \dots, n_d)$ with the integer coordinates n_i. Continuous-time branching random walks (BRWs) on $\mathbb{Z}^d$ have been widely discussed in a series see, for example, [1, 2, 9–12, 14].

Until recently, BRWs with one branching source, i.e. the source of reproduction and death of particles, have been the main object of interest in investigations of such models. Various methods have been applied for analyses of this case, and an approach based on representation of the evolutionary equations for moments of numbers of particles as equations in Banach spaces [1, 2, 12, 14] has found to be rather productive.

The BRW with one branching source was investigated in [3, 12], where the conditions for exponential growth of the numbers of particles were obtained for an arbitrary point as well as for the entire lattice. Such a character of growth of the number of particles is determined by the existence of an isolated positive eigenvalue in the spectrum of the operator of the mean number of particles. In this model, the evolutionary operator can be represented as the sum of a bounded self-adjoint operator and a completely continuous one, which gives opportunity to obtain the

---

E. Antonenko • E. Yarovaya (✉)
Department of Probability Theory, Lomonosov Moscow State University, Leninskie Gory, Main Building, Moscow, Russia
e-mail: eka.antonenko@gmail.com; yarovaya@mech.math.msu.su

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_3

necessary and sufficient conditions for existence of an isolated positive eigenvalue in the spectrum of the evolutionary operator.

We consider the model of continuous-time BRW on $\mathbb{Z}^d$ with the finite number N of the branching sources located at some lattice points $x_1, x_2, \dots, x_N$. It is assumed that the particles evolve independently of each other and of the prehistory. Each particle walks on the lattice until it reaches a branching source where its behavior changes. We assume that death and reproduction of a particle happen without breaking the symmetry of the underlying random walk. In this case the behaviour of the mean number of particles can be described in terms of the evolutionary operator of a special type [15], which is a perturbation of the generator $\mathscr{A}$ of a symmetric random walk. In the case of equal intensities of sources this operator has the form

$$\mathscr{H}_\beta = \mathscr{A} + \beta \sum_{i=1}^{N} \delta_{x_i} \delta_{x_i}^T, \quad x_i \in \mathbb{Z}^d, \tag{3.1}$$

where $\mathscr{A} : l^p(\mathbb{Z}^d) \to l^p(\mathbb{Z}^d)$, $p \in [1, \infty]$, is a symmetric operator and $\delta_x = \delta_x(\cdot)$ denotes a column vector on the lattice taking the value one at the point x and zero otherwise. General analysis of this operator was first done in [15].

In [16] it is shown how the operators of type (3.1) appear in BRW models and is demonstrated that the structure of its spectrum determines the asymptotic behaviour of the numbers of particles. For the analysis of the evolutionary equations for the mean number of particles, in [16] there was used the technique of differential equations in Banach spaces. In [15] it is shown that $\mathscr{H}_\beta$ is a linear bounded operator in every space $l^p(\mathbb{Z}^d)$, $p \in [1, \infty]$. All points of its spectrum outside the circle $C = \{z \in \mathbb{C} : |z - a(0)| \leq |a(0)|\}$ with $a(0) = \delta_0^T \mathscr{A} \delta_0$ may be only eigenvalues of finite multiplicity. This statement allowed to propose a general method for obtaining a finite set of equations defining conditions of the existence of isolated positive eigenvalues in the spectrum of the operator $\mathscr{H}_\beta$ lying outside C. Corresponding conditions for BRWs with two sources of branching were obtained in explicit form in [15].

In [15] it is shown, in particular, that the perturbation of the form $\beta \sum_{i=1}^{N} \delta_{x_i} \delta_{x_i}^T$ of the operator $\mathscr{A}$ may result in the emergence of positive eigenvalues of the operator $\mathscr{H}_\beta$ and the multiplicity of each of them does not exceed N. However, the number of arising eigenvalues of $\mathscr{H}_\beta$ in [15] was not found. In the present paper we strengthen this result by showing in Theorem 3.3 that the maximum eigenvalue of the operator $\mathscr{H}_\beta$ is of unit multiplicity, that is simple, and the general multiplicity of all eigenvalues does not exceed N. This implies, in particular, that in fact the multiplicity of each eigenvalue of the operator $\mathscr{H}_\beta$ does not exceed $N - 1$. Example 3.1 demonstrates that the greatest possible multiplicity $N - 1$ of a positive eigenvalue of the operator $\mathscr{H}_\beta$ can actually be reached when $N > 1$.

The structure of the present paper is as follows. In Sect. 3.2 a formal description of BRW with N sources is reminded. In Sect. 3.3 the main results are formulated and the example about influence of "symmetry" of the sources disposition on appearance of coinciding eigenvalues in the spectrum of operator (3.1) is given. Section 3.4 contains proofs of the formulated theorems.

3.2 BRW with Several Sources

The evolution of the system of particles in BRW on $\mathbb{Z}^d$ is defined by the number of particles $\mu_t(y)$ at moment t at each point $y \in \mathbb{Z}^d$ assuming that the system contains only one particle disposed at some point $x \in \mathbb{Z}^d$ at $t = 0$, i.e. $\mu_0(y) = \delta_x(y)$. Thus, the total number of particles on $\mathbb{Z}^d$ satisfies the equation $\mu_t = \sum_{y \in \mathbb{Z}^d} \mu_t(y)$. The transition probability of the random walk in the BRW is denoted by $p(t, x, y)$. Let E_x be the expectation of the total number of particles on the condition that $\mu_0(\cdot) = \delta_x(\cdot)$. Then, the moments obey $m_n(t, x, y) := \mathsf{E}_x \mu_t^n(y)$ and $m_n(t, x) := \mathsf{E}_x \mu_t^n$, $n \in \mathbb{N}$.

Random walk is specified by a matrix $A = (a(x, y))_{x,y \in \mathbb{Z}^d}$ of transition intensities, where $a(x, y) = a(0, x-y) = a(x-y)$ for all x and y. Thus, the transition intensities are spatially homogeneous and the matrix A is symmetric. The law of walk is described in terms of the function $a(z)$, $z \in \mathbb{Z}^d$, where $a(0) < 0$, $a(z) \geq 0$ when $z \neq 0$ and $a(z) \equiv a(-z)$. We assume that $\sum_{z \in \mathbb{Z}^d} a(z) = 0$ and

$$\sum_{z \in \mathbb{Z}^d} |z|^2 a(z) < \infty, \tag{3.2}$$

where $|z|$ is the Euclidian norm of the vector z. We also assume that the matrix A is irreducible, i.e. for all $z \in \mathbb{Z}^d$ there exists a set of vectors $z_1, z_2, \ldots, z_k \in \mathbb{Z}^d$ such that $z = \sum\limits_{i=1}^{k} z_i$ and $a(z_i) \neq 0$ for $i = 1, 2, \ldots, k$.

By $p(t, x, y)$ we denote the transition probability of a random walk. This function is implicitly determined by the transition intensities $a(x, y)$ (see, for example, [6, 12]). Then, Green's function of the operator $\mathscr{A}$ can be represented as a Laplace transform of the transition probability $p(t, x, y)$:

$$G_\lambda(x, y) := \int\limits_0^\infty e^{-\lambda t} p(t, x, y)\, dt, \qquad \lambda \geq 0.$$

The analysis of BRWs essentially depends on whether the value $G_0 = G_0(0, 0)$ is finite or not. If the condition (3.2) (finiteness of the variance of jumps) is satisfied then $G_0 = \infty$ when $d = 1, 2$, and $G_0 < \infty$ when $d \geq 3$ (see, for example [12]).

We use the function b_n, $n \geq 0$, where $b_n \geq 0$ for $n \neq 1$, $b_1 \leq 0$, and $\sum_n b_n = 0$, to describe branching at a source. Branching occurs at a finite number of sources, $x_1, \ldots, x_N$, and is given by the infinitesimal generating function $f(u) = \sum_{n=0}^{\infty} b_n u^n$ such that $\beta_r = f^{(r)}(1) < \infty$ for all $r \in \mathbb{N}$. The quantity $\beta_1 = f'(1)$ characterizes the intensity of a source and is denoted further by β. The sojourn time of a particle at every source is distributed exponentially with the parameter $-(a(0)+b_1)$, see [13].

The analysis of the BRW model with one branching source in [2, 3, 12, 14] showed that the asymptotic behaviour of the mean number of particles at arbitrary point as well as on the entire lattice is determined by the structure of the spectrum of the linear operator (3.1) when $N = 1$. Here the bounded self-adjoint operator $\mathscr{A}$ in Hilbert space $l^2(\mathbb{Z}^d)$ is a generator of random walk, and $\beta\Delta_{x_1}$ specifies the mechanism of branching at the source x_1. Let us note that the operator $\mathscr{A}$ is generated by the matrix A of transition intensities. This model has been generalized in [15], in particular to the case of N sources.

The transition probability $p(t,\cdot,y)$ is treated as a function $p(t)$ in $l^2(\mathbb{Z}^d)$ depending on time t and the parameter y. Then according to [12, 15] we can rewrite the evolution equation as the following differential equation in space $l^2(\mathbb{Z}^d)$:

$$\frac{dp}{dt} = \mathscr{A}p, \qquad p(0) = \delta_y.$$

Here the operator $\mathscr{A}$ acts as

$$(Au)(z) := \sum_{z'\in\mathbb{Z}^d} a(z-z')u(z').$$

In the same way we can obtain the differential equation in space $l^2(\mathbb{Z}^d)$ for the expectation $m_1(t,\cdot,y)$ which can be considered as a function $m_1(t)$ in $l^2(\mathbb{Z}^d)$:

$$\frac{dm_1}{dt} = \mathscr{H}_\beta m_1, \quad m_1(0) = \delta_y. \tag{3.3}$$

Formally, this equation holds for $m_1(t) = m_1(t,\cdot)$ on the condition that $m_1(0) = 1$ in space $l^\infty(\mathbb{Z}^d)$.

It follows from the general theory of linear differential equations in Banach spaces (see, for example, [4]) that the investigation of behaviour of solutions of the Eq. (3.3) can be reduced to the analysis of the spectrum of the linear operators in the right-hand sides of the corresponding equations. Spectral analysis of the operator $\mathscr{H}_\beta$ of type (3.1) was done in [15].

3.3 Spatial Configuration of Sources of Branching

We denote by β_c the value of intensity of sources uniquely defined by the condition that the spectrum of the operator $\mathscr{H}_\beta$ contains positive eigenvalues if $\beta > \beta_c$.

Let us formulate the main results to be proved below.

Theorem 3.1 *If $d = 1$ or $d = 2$, then $\beta_c = 0$; if $d \ge 3$, then $\beta_c = (G_0)^{-1}$ in the case $N = 1$, and $0 < \beta_c < (G_0)^{-1}$ in the case $N \ge 2$.*

Definition 3.1 If there exists $\varepsilon_0 > 0$ such that the operator $\mathscr{H}_\beta$ has a simple positive eigenvalue $\lambda(\beta)$ when $\beta \in (\beta_c, \beta_c + \varepsilon_0)$ and this eigenvalue satisfies a condition $\lambda(\beta) \to 0$ as $\beta \downarrow \beta_c$, then we call supercritical BRW *weakly supercritical* when β is close to β_c.

From the definition above, the question arises *whether or not every supercritical BRW is weakly supercritical*. The following theorem gives the affirmative answer to this question.

Theorem 3.2 *Every supercritical BRW is weakly supercritical as* $\beta \downarrow \beta_c$.

This theorem follows from the following stronger statement.

Theorem 3.3 *Let* $N \geq 2$*, then the operator* $\mathscr{H}_\beta$ *may have no more than N positive eigenvalues* $\lambda_i(\beta)$ *of finite multiplicity when* $\beta > \beta_c$*, and*

$$\lambda_0(\beta) > \lambda_1(\beta) \geq \cdots \geq \lambda_{N-1}(\beta) > 0.$$

Here the principal eigenvalue $\lambda_0(\beta)$ *has unit multiplicity. Besides there is a value* β_{c_1} *such that for* $\beta \in (\beta_c, \beta_{c_1})$ *the operator* $\mathscr{H}_\beta$ *has a single positive eigenvalue,* $\lambda_0(\beta)$.

Corollary 3.1 *Under the conditions of Theorem 3.3 the multiplicity of each of the eigenvalues* $\lambda_1(\beta), \ldots, \lambda_{N-1}(\beta)$ *does not exceed* $N - 1$.

We can use the following theorem, which was established in different terms in [15, Theorem 6], for estimating the eigenvalues

$$\lambda_0(\beta), \lambda_1(\beta), \ldots, \lambda_{N-1}(\beta).$$

Theorem 3.4 *The eigenvalue* λ *belongs to the discrete spectrum of the operator* $\mathscr{H}_\beta$ *in the only case when the system of linear equations*

$$V_i - \beta \sum_{j=1}^{N} G_\lambda(x_i - x_j) V_j = 0, \quad i = 1, \ldots, N \tag{3.4}$$

in variables $\{V_i\}_{i=1}^N$ *has a non-trivial solution.*

According to Theorem 3.3 the principal eigenvalue $\lambda_0(\beta)$ is always simple. Some of the eigenvalues $\lambda_1(\beta), \ldots, \lambda_{N-1}(\beta)$ of the operator $\mathscr{H}_\beta$ may actually coincide (i.e. they may have the multiplicity greater than one) and this situation is possible even in the case of an arbitrary finite number of sources (of equal intensity). As is shown in the following Example 3.1 this situation may occur if there is a certain "symmetry" of the spatial configuration of the sources $x_1, x_2, \ldots, x_N$.

In Example 3.1 we assume that the function of transition probabilities is symmetric that is its values do not change at any permutation of arguments. In particular, a function of a vector variable z is symmetric if its values are the same at any permutation of coordinates of vector z. Let us present a statement to be use further for “integrable” models, where critical values β_{c_i}, $i \geq 2$, and equations for estimating λ_i from Theorem 3.3 can be found explicitly.

Theorem 3.5 *If the function of transition probabilities $a(z)$ is symmetric, then the function $G_\lambda(z)$ is also symmetric.*

Example 3.1 Let $x_1, \dots, x_N$, when $N \geq 2$, be the vertices of a regular simplex (i.e. lengths of the simplex edges are equal). For example,

$$x_1 = \{1, 0, \dots, 0\},\ x_2 = \{0, 1, \dots 0\},\ \dots,\ x_N = \{0, 0, \dots, 1\}.$$

The existence of a non-trivial solution of the linear equation (3.4) for some β is equivalent to the resolvability of the equation

$$\det\left(\varGamma(\lambda) - \tfrac{1}{\beta} I\right) = 0, \tag{3.5}$$

where

$$\varGamma(\lambda) = [\varGamma_{ij}(\lambda)] \tag{3.6}$$

is the matrix with elements $\varGamma_{ij}(\lambda) = G_\lambda(x_i, x_j)$.

Since the random walk by assumption is symmetric and homogeneous then

$$G_\lambda(x_i, x_j) = G_\lambda(0, x_i - x_j) = G_\lambda(0, x_j - x_i) = G_\lambda(x_j - x_i).$$

From the definition of the function $G_\lambda(u, v) \equiv G_\lambda(u - v)$ it follows that all the values $G_\lambda(x_j - x_i)$ coincide with each other when $i \neq j$ and hence they coincide with $G_\lambda(x_1 - x_2) = G_\lambda(z_*)$ (for simplicity we denote $z_* = x_1 - x_2$). So,

$$G_\lambda(x_j - x_i) \equiv G_\lambda(x_1 - x_2) = G_\lambda(z_*) \quad \text{for all} \quad i \neq j. \tag{3.7}$$

For the values $G_\lambda(x_i, x_i)$ we have

$$G_\lambda(x_i, x_i) \equiv G_\lambda(x_i - x_i) = G_\lambda(0) = G_\lambda.$$

Thus we can rewrite Eq. (3.5) as

$$\det \begin{bmatrix} G_\lambda - \frac{1}{\beta} & \cdots & G_\lambda(z_*) \\ G_\lambda(z_*) & \cdots & G_\lambda(z_*) \\ \cdots & \cdots & \cdots \\ G_\lambda(z_*) & \cdots & G_\lambda - \frac{1}{\beta} \end{bmatrix} = 0.$$

Using standard linear transforms we rewrite the last determinantal equation as

$$\left(G_\lambda - G_\lambda(z_*) - \frac{1}{\beta}\right)^{N-1} \det \begin{bmatrix} G_\lambda - \frac{1}{\beta} + (N-1)G_\lambda(z_*) & \cdots & G_\lambda(z_*) \\ 0 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & -1 \end{bmatrix} = 0.$$

This is equivalent to the equation

$$\left(G_\lambda + (N-1)G_\lambda(z_*) - \frac{1}{\beta}\right)\left(G_\lambda - G_\lambda(z_*) - \frac{1}{\beta}\right)^{N-1} = 0.$$

From this last equation it is seen that β_c and β_{c_1} can be calculated explicitly:

$$\beta_c = (G_0 + (N-1)G_0(z_*))^{-1}, \quad \beta_{c_1} = (G_0 - G_0(z_*))^{-1}. \tag{3.8}$$

Remark 3.1 Under the conditions of Example 3.1 according to (3.7) and (3.8) the quantity β_{c_1} depends on the norm $|z_*|$ of the vector z_* (i.e. on the distance between the sources) and does not depend on the number of the sources N, that is $\beta_{c_1} = \beta_c(|z_*|) > 0$. At the same time the quantity β_c depends not only on the distance between the sources but also on the number of the sources N, that is $\beta_c = \beta_c(|z_*|, N)$, and in such a way that $\beta_c(|z_*|, N) \to 0$ as $N \to \infty$ when z_* is fixed. Moreover, $\beta_c(|z_*|, N) \equiv 0$ when $d = 1$ and $d = 2$.

3.4 Proofs of Theorems

To prove Theorems 3.1 and 3.2 we will use the following lemmas.

Lemma 3.1 *The quantity $\lambda > 0$ is an eigenvalue of the operator $\mathscr{H}_\beta$ if and only if at least one of the equations*

$$\gamma_i(\lambda)\beta = 1, \quad i = 0, \ldots, N-1, \tag{3.9}$$

holds, where $\gamma_i(\lambda)$ are eigenvalues of the matrix $\Gamma(\lambda)$ given by the Eq. (3.6).

Proof The quantity $\lambda > 0$ is an eigenvalue of the operator $\mathscr{H}_\beta = \mathscr{A} + \beta \sum_{i=1}^N \delta_{x_i}\delta_{x_i}^T$ if and only if the following equation holds for some vector $h \neq 0$:

$$\mathscr{A}h + \beta \sum_{i=1}^N \delta_{x_i}\delta_{x_i}^T h = \lambda h.$$

Let $R_\lambda = (\mathscr{A} - \lambda I)^{-1}$ be the resolvent of the operator $\mathscr{A}$. By applying R_λ to both sides of the last equation we obtain

$$h + \beta \sum_{i=1}^{N} R_\lambda \delta_{x_i} \delta_{x_i}^T h = 0.$$

Since $\delta_x \delta_x^T h = \delta_x(\delta_x, h)$, then

$$h + \beta \sum_{i=1}^{N} (\delta_{x_i}, h) R_\lambda \delta_{x_i} = 0.$$

Let us scalar left-multiply the last equation by δ_{x_k}:

$$(\delta_{x_k}, h) + \sum_{i=1}^{n} \beta(\delta_{x_i}, h)(\delta_{x_k}, R_\lambda \delta_{x_i}) = 0, \quad k = 1, \dots, n.$$

By denoting $U_k = (\delta_{x_k}, h)$ we then obtain

$$U_k + \sum_{i=1}^{n} \beta U_i (\delta_{x_k}, R_\lambda \delta_{x_i}) = 0, \quad k = 1, \dots, n. \tag{3.10}$$

Thus, the initial equation has a nonzero solution h if and only if the determinant of the matrix of the obtained linear system is equal to zero. Now we notice that

$$(\delta_y, R_\lambda \delta_x) = -\frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{i(\theta, y-x)}}{\lambda - \phi(\theta)} d\theta,$$

where $\phi(\theta) = \sum_{z \in \mathbb{Z}^d} a(z) e^{i(\theta, z)}$ with $\theta \in [-\pi, \pi]^d$ is the Fourier transform of the function of transition probabilities $a(z)$. The right-hand side of the equation can be represented [12] in terms of Green's function:

$$G_\lambda(x, y) := \int_0^\infty e^{-\lambda t} p(t, x, y) dt = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{i(\theta, y-x)}}{\lambda - \phi(\theta)} d\theta. \tag{3.11}$$

Hence $(\delta_y, R_\lambda \delta_x) = -G_\lambda(x, y)$. It implies that the condition of vanishing of the determinant of the linear system (3.10) can be rewritten as $\det(\beta \Gamma(\lambda) - I) = 0$, which is equivalent to Eq. (3.5) when $\beta \neq 0$.

Recalling that the eigenvalues of the matrix $\Gamma(\lambda)$ are denoted by $\gamma_i(\lambda)$, where $i = 0, \dots, N-1$, we obtain that (3.5) holds for some β and λ if and only if (3.9) is true. The lemma is proved. □

Lemma 3.2 *Each of the functions $\gamma_i(\lambda)$ is strictly decreasing when $\lambda \geq 0$. Moreover, the total number of solutions $\lambda_i(\beta)$ of the equations $\gamma_i(\lambda)\beta = 1$,*

$i = 0, \dots, N-1$, does not exceed N, that is the number of eigenvalues of the operator (3.1) *does not exceed N.*

We need the following auxiliary statement to prove Lemma 3.2.

Lemma 3.3 *Let $Q = (q_{ij})$ be a matrix with elements*

$$q_{ij} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} q(\theta) e^{i(\theta, x_i - x_j)} d\theta,$$

where $x_1, x_2, \dots, x_N$ is a set of linearly independent vectors, and $q(\theta) \geq q_* > 0$ is an even function summable on $[-\pi, \pi]^d$. Then Q is a real, symmetric and positive-definite matrix satisfying $(Qz, z) \geq q_*(z, z)$.

Proof The matrix Q is real and symmetric since the function $q(\theta)$ is even and then

$$q_{ij} = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} q(\theta) \cos(\theta, x_i - x_j) d\theta.$$

Thus, we need to prove only that the matrix Q is positive-definite. By definition, $(Qz, z) = \sum_{i,j=1}^{N} q_{ij} z_i z_j$, where $z = (z_1, z_2, \dots, z_N)$. Then

$$\begin{aligned}
(Qz, z) &= \frac{1}{(2\pi)^d} \sum_{i,j=1}^{N} \int_{[-\pi,\pi]^d} q(\theta) e^{i(\theta, x_i - x_j)} z_i z_j d\theta \\
&= \frac{1}{(2\pi)^d} \sum_{i=1}^{N} \sum_{j=1}^{N} \int_{[-\pi,\pi]^d} q(\theta) \left(e^{i(\theta, x_i)} z_i\right) \left(e^{-i(\theta, x_j)} z_j\right) d\theta \\
&= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} q(\theta) \sum_{i=1}^{N} \sum_{j=1}^{N} \left(e^{i(\theta, x_i)} z_i\right) \left(e^{-i(\theta, x_j)} z_j\right) d\theta \\
&= \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} q(\theta) \left| e^{i(\theta, x_1)} z_1 + \dots + e^{i(\theta, x_N)} z_N \right|^2 d\theta \geq 0.
\end{aligned}$$

Since $q(\theta) \geq q_*$, then

$$\begin{aligned}
(Qz, z) &\geq \frac{q_*}{(2\pi)^d} \int_{[-\pi,\pi]^d} \left| e^{i(\theta, x_1)} z_1 + \dots + e^{i(\theta, x_N)} z_N \right|^2 d\theta \\
&= \frac{q_*}{(2\pi)^d} \int_{[-\pi,\pi]^d} \left(z_1^2 + \dots + z_N^2 + \sum_{i \neq j} \left(e^{i(\theta, x_i)} z_i\right) \left(e^{-i(\theta, x_j)} z_j\right) \right) d\theta.
\end{aligned}$$

The integral of the summands $\left(e^{i(\theta, x_i)} z_i\right) \left(e^{-i(\theta, x_j)} z_j\right)$ vanishes since $x_i \neq x_j$ when $i \neq j$. It implies that the integral in the right-hand side of the equation can be found explicitly and is equal to $q_*(z_1^2 + \ldots + z_N^2)$. So, $(Qz, z) \geq q_*(z, z)$. The lemma is proved. □

Proof of Lemma 3.2 Let the eigenvalues $\gamma_i(\lambda)$ of the matrix (3.6) be arranged in decreasing order:

$$0 \leq \gamma_{N-1}(\lambda) \leq \ldots \leq \gamma_2(\lambda) < \gamma_0(\lambda).$$

Let us consider a matrix

$$\Gamma(\lambda_1, \lambda_2) := \Gamma(\lambda_1) - \Gamma(\lambda_2).$$

Taking

$$q(\theta) = \frac{1}{\lambda - \phi(\theta)}$$

we obtain for $\lambda > 0$ that

$$q(\theta) \geq \frac{1}{\lambda + s} > 0,$$

where $s = \max_{\theta \in [-\pi,\pi]^d} \{-\phi(\theta)\} > 0$. Hence, by Lemma 3.3 for each $\lambda > 0$ the matrix $\Gamma(\lambda)$ defined by (3.6) and (3.11) is real, symmetric and positive-definite.

From (3.6) and (3.11) we obtain also that the elements of the matrix $\Gamma(\lambda_1, \lambda_2)$ are as follows

$$\Gamma_{ij}(\lambda_1, \lambda_2) = (\lambda_2 - \lambda_1) \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \frac{e^{i(\theta, x_i - x_j)}}{(\lambda_1 - \phi(\theta))(\lambda_2 - \phi(\theta))} d\theta.$$

For the continuous function

$$q(\theta) = \frac{1}{(\lambda_1 - \phi(\theta))(\lambda_2 - \phi(\theta))}$$

we have the lower bound

$$q(\theta) \geq q_*(\lambda_1, \lambda_2) := \frac{1}{(\lambda_1 + s)(\lambda_2 + s)} > 0.$$

Hence, again by Lemma 3.3 the matrix $\Gamma(\lambda_1, \lambda_2)$ is self-adjoint and positive-definite when $\lambda_2 > \lambda_1$. In this case the Lidskii theorem [8, Theorem 6.10] implies, for all $i = 0, \ldots, N-1$, the inequalities

$$\gamma_i(\lambda_1) - \gamma_i(\lambda_2) \geq q_*(\lambda_1, \lambda_2) > 0,$$

since the minimal eigenvalue of the matrix $\Gamma(\lambda_1, \lambda_2)$ has the lower bound equal to $q_*(\lambda_1, \lambda_2)$. So, $\gamma_i(\lambda_1) > \gamma_i(\lambda_2)$ when $\lambda_2 > \lambda_1$, that is the function $\gamma_i(\lambda)$ is strictly decreasing with respect to λ.

Since the functions $\gamma_i(\lambda)$ are strictly decreasing then each of the equations $\gamma_i(\lambda)\beta = 1$, where $i = 0, \ldots, N-1$, for each β has no more than one solution (the eigenvalue of the operator $\mathcal{H}_\beta$). So, the total amount of the eigenvalues of the operator $\mathcal{H}_\beta$ does not exceed N. The lemma is proved. □

Proof of Theorem 3.1 In the integral representation (3.11), the function $\phi(\theta)$ satisfies $c\|\theta\|^2 \leqslant |\phi(\theta)| \leqslant C\|\theta\|^2$ for some nonzero real constants c and C [12]. Hence the convergence of the integral in the integral representation of the Green's function $G_\lambda(x, y)$ as $\lambda \to 0$ is equivalent [12] to the convergence of the integral

$$\int \frac{r^{d-1}}{r^2} dr$$

in a neighbourhood of 0. This last integral converges when $d \geqslant 3$ and diverges when $d = 1$ and $d = 2$. If $G_\lambda(x, y) \to \infty$ as $\lambda \to 0$ then $\|\Gamma(\lambda)\| \to \infty$ and the principal eigenvalue of matrix $\Gamma(\lambda)$ tends to infinity as $\lambda \to 0$, $\gamma_0(\lambda) \to \infty$. Hence in this case for all $\beta > 0$ the equation $\gamma_0(\lambda)\beta = 1$ has a solution (with respect to λ) and by definition of β_c we have that $\beta_c = 0$.

Let now $G_0(0, 0) < \infty$, then $G_0(x, y) < \infty$ for all x and y. So, in this case $\|\Gamma(0)\| < \infty$ and, moreover, $\Gamma(\lambda) \to \Gamma(0)$ as $\lambda \to 0$. Then there exists $\gamma_* < \infty$ such that $\gamma_0(\lambda) \leq \gamma_* < \infty$ for all λ. In this case the equation $\gamma_0(\lambda)\beta = 1$ does not have solutions (with respect to λ) as $\beta \to 0$. By Lemma 3.1 in this case the operator $\mathscr{H}_\beta$ does not have eigenvalues when β is small, i.e. $\beta_c > 0$.

Let us prove the upper bound for β_c when $d \geqslant 3$. Let β be an arbitrary value of the parameter such that the operator $\mathscr{H}_\beta$ has a positive eigenvalue λ. Then by Lemma 3.1 we have $\gamma_0(\lambda)\beta = 1$, and hence

$$\beta = \frac{1}{\gamma_0(\lambda)}. \tag{3.12}$$

The Perron–Frobenius theorem [5] implies that the principal eigenvalue $\gamma_0(\lambda)$ of the matrix $\Gamma(\lambda)$ has a corresponding eigenvector $x(\lambda)$ with all positive coordinates. The matrix $\Gamma(\lambda)$ can be represented as

$$\Gamma(\lambda) = G_\lambda(0, 0)I + B(\lambda),$$

where $B(\lambda) = \Gamma(\lambda) - G_\lambda(0, 0)I$. All elements of the matrix $B(\lambda)$ are non-negative while its off-diagonal elements are positive when $N \geq 2$. In this case by definition of the eigenvector $x(\lambda)$ the following equalities hold:

$$0 = \Gamma(\lambda)x(\lambda) - \gamma_0(\lambda)x(\lambda) = (G_\lambda(0, 0) - \gamma_0(\lambda))x(\lambda) + B(\lambda)x(\lambda).$$

Hence, when $\lambda = 0$,

$$0 = \Gamma(0)x(0) - \gamma_0(0)x(0) = (G_0(0,0) - \gamma_0(0))x(0) + B(0)x(0).$$

The vector $x(0)$ has positive coordinates and therefore the vector $B(0)x(0)$ also has positive coordinates. So, the last equation holds only if $G_0(0,0) - \gamma_0(0) < 0$. Hence, by (3.12) we have

$$\beta_c = \frac{1}{\gamma_0(0)} < \frac{1}{G_0(0,0)} = \frac{1}{G_0}.$$

For $N = 1$ the critical value β_c can be found from the equation $\beta_c G_0 = 1$ and equals $\beta_c = \frac{1}{G_0}$. □

Proof of Theorem 3.3 By Lemma 3.1 the eigenvalues of the operator $\mathscr{H}_\beta$ satisfy the Eqs. (3.9). The quantities $\gamma_i(\lambda)$ are the eigenvalues of the positive and symmetric matrix $\Gamma(\lambda)$, and (as is shown in [9] and also follows from Lemma 3.3) this matrix is the Gramian matrix for some appropriate scalar product and hence is positive-definite. In this case all eigenvalues $\gamma_i(\lambda)$ are real and non-negative (and can be arranged in ascending order):

$$0 \le \gamma_{N-1}(\lambda) \le \ldots \le \gamma_1(\lambda) \le \gamma_0(\lambda). \tag{3.13}$$

From (3.11) it follows that elements of the matrix $\Gamma(\lambda)$ tend to zero as $\lambda \to \infty$ and so $\gamma_i(\lambda) \to 0$ as $\lambda \to \infty$ for all $i = 0, 1, \ldots, N-1$. By Rellich theorem [8, Chap. 2, Theorem 6.8] all functions $\gamma_i(\lambda)$ are piecewise smooth when $\lambda \ge 0$.

Using the representation (3.11) we notice that elements of the matrix $\Gamma(\lambda)$ are strictly positive when $\lambda > 0$ and then by the Perron–Frobenius theorem [5] we have that the principal eigenvalue $\gamma_0(\lambda)$ of the matrix $\Gamma(\lambda)$ has the unit multiplicity and strictly exceeds other eigenvalues, i.e. the last of the inequalities (3.13) is strict:

$$0 \le \gamma_{N-1}(\lambda) \le \ldots \le \gamma_1(\lambda) < \gamma_0(\lambda). \tag{3.14}$$

From the representations (3.11) it follows that the matrix $\Gamma(\lambda)$ is continuous for all values $\lambda > 0$. Behaviour of this matrix can differ as $\lambda \to 0$ and further proof of the theorem depends on this behaviour.

The matrix $\Gamma(\lambda)$ may tend to some finite limit matrix $\Gamma(0)$ with strictly positive elements as $\lambda \to 0$. In this case the eigenvalues $\gamma_i(\lambda)$ behave as it is shown on Fig. 3.1 and the difference between the values $\gamma_0(0) = \frac{1}{\beta_c}$ and $\gamma_1(0) = \frac{1}{\beta_{c_1}}$ can be estimated by, for example, the Hopf theorem [7, p. 592]. This is true for $d \ge 3$.

Another case is when at least one element of the matrix $\Gamma(\lambda)$ has infinite limit as $\lambda \to 0$. This is possible only when $d = 1$ or $d = 2$. In this case $\|\Gamma(\lambda)\| \to \infty$ as $\lambda \to 0$ and hence $\gamma_0(\lambda) \to \infty$ as $\lambda \to 0$. This happens only for $\beta_c = 0$. We can suppose that $\gamma_1(\lambda) \to \infty$ as $\lambda \to 0$, i.e. $\beta_{c_1} = 0$. However, it follows from the

Fig. 3.1 Plots of the functions $\gamma_i(\lambda)$ when $\lim_{\lambda\to\infty} \gamma_0(\lambda) < \infty$

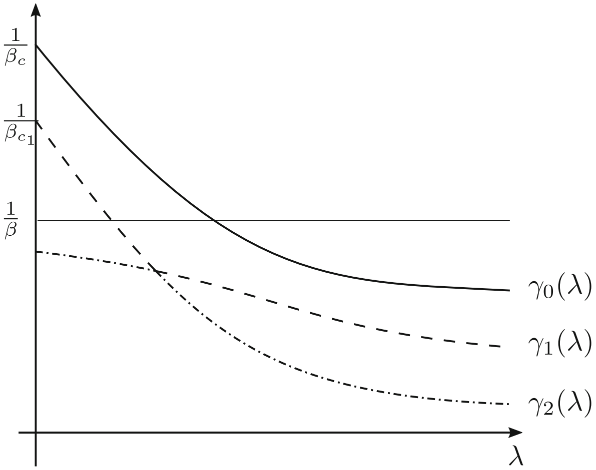
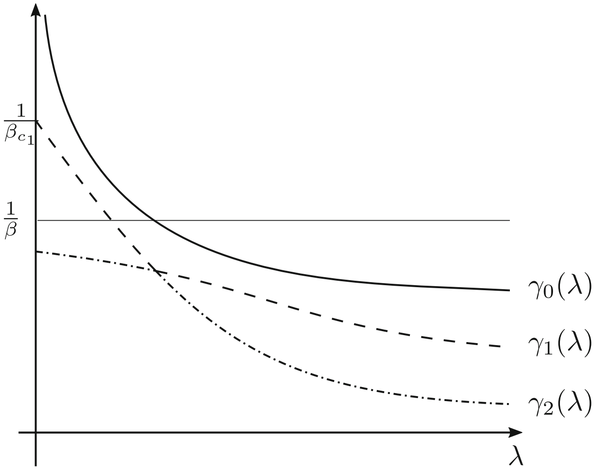


Fig. 3.2 Plots of the functions $\gamma_i(\lambda)$ when $\lim_{\lambda\to\infty} \gamma_0(\lambda) = \infty$, but $\lim_{\lambda\to\infty} \gamma_1(\lambda) < \infty$



Theorem 2.1.1 from [12] that this situation is impossible and $\gamma_1(\lambda)$ always has finite limit as $\lambda \to 0$, i.e. $\beta_{c_1} > 0$. This situation is illustrated on Fig. 3.2.

By Lemma 3.1 the eigenvalues λ_i of the operator $\mathscr{H}_\beta$ are solutions of the Eqs. (3.9), see Fig. 3.1. However, since by Lemma 3.2 every function $\gamma_i(\lambda)$ is strictly decreasing when $\lambda \geq 0$, then the total number of the solutions $\lambda_i(\beta)$ of these equations does not exceed N.

From inequalities (3.14) it follows that if the operator $\mathscr{H}_\beta$ has positive eigenvalues for some fixed β, then the maximal of them is $\lambda_0 = \lambda_0(\beta)$ which is a solution of the equation

$$\gamma_0(\lambda_0) = \frac{1}{\beta}, \tag{3.15}$$

is simple and strictly exceeds others. The minimum value of the solution λ_0 of Eq. (3.15) is $\lambda_0 = 0$. The corresponding value β_c of the parameter β is critical. Since the function $\gamma_0(\lambda)$ is strictly decreasing then the eigenvalue $\lambda_0 = \lambda_0(\beta)$ increases with the increase of parameter β. The theorem is proved. □

Proof of Theorem 3.5 Let us recall that Green's function $G_\lambda(z)$ can be represented [12] in the form (3.11), where $\phi(\theta)$ is the Fourier transform of the function of transition probabilities $a(z)$ and is defined by $\phi(\theta) = \sum_{z\in\mathbb{Z}^d} a(z)e^{i(\theta,z)}$, $\theta \in [-\pi,\pi]^d$. To prove the theorem it suffices to demonstrate that $G_\lambda(z) = G_\lambda(\mathbf{R}z)$ for every permutation matrix $\mathbf{R}$ (i.e. all rows and columns of this matrix have the only one nonzero element equal to one). So,

$$\begin{aligned}\phi(\mathbf{R}\theta) &= \sum_{z\in\mathbb{Z}^d} a(z)e^{i(\mathbf{R}\theta,z)} = \sum_{z\in\mathbb{Z}^d} a(z)e^{i(\theta,\mathbf{R}^*z)}\\ &= \sum_{z'\in\mathbb{Z}^d} a((\mathbf{R}^*)^{-1}z')e^{i(\theta,z')} = \sum_{z'\in\mathbb{Z}^d} a(z')e^{i(\theta,z')} = \phi(\theta),\end{aligned}$$

where the equality $a((\mathbf{R}^*)^{-1}z') = a(z')$ holds for all $z' \in \mathbb{Z}^d$ since the function $a(z)$ is symmetric. Hence the function $\phi(\theta)$ is also symmetric. Further,

$$\begin{aligned}G_\lambda(\mathbf{R}z) &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{e^{i(\theta,\mathbf{R}z)}}{\lambda-\phi(\theta)}d\theta = \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{e^{i(\mathbf{R}^*\theta,z)}}{\lambda-\phi(\theta)}d\theta\\ &= \frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d}\frac{e^{i(\theta,z)}}{\lambda-\phi((\mathbf{R}^*)^{-1}\theta)}d\theta = G_\lambda(z)\end{aligned}$$

for every permutation matrix $\mathbf{R}$, where the equality $\phi(\theta) = \phi((\mathbf{R}^*)^{-1}\theta)$ again holds for all $\theta \in [-\pi,\pi]^d$ since the function $\phi(\theta)$ is symmetric. Thus, the function $G_\lambda(z)$ is also symmetric. □

Acknowledgements This study was performed at Lomonosov Moscow state University and at Steklov Mathematical Institute, Russian Academy of Sciences. The work was supported by the Russian Science Foundation, project no. 14-21-00162.

References

1. Albeverio, S., Bogachev, L.: Branching random walk in a catalytic medium. Basic Equ. (2000). doi:10.1023/A:1009818620550
2. Albeverio, S., Bogachev, L., Yarovaya, E.: Asymptotics of branching symmetric random walk on the lattice with a single source. C. R. Acad. Sci. Paris Sér. I Math. (1998). doi:10.1016/S0764-4442(98)80125-0
3. Bogachev, L., Yarovaya, E.: A limit theorem for a supercritical branching random walk on $\mathbb{Z}^d$ with a single source. Russ. Math. Surv. **53**(5), 1086–1088 (1998)

4. Daleckii, J., Krein, M.: Stability of Solutions of Differential Equations in Banach Space. AMS, Providence (1974)
5. Gantmacher, F.: Theory of Matrices. AMS Chelsea Publishing, New York (1959)
6. Gikhman, I., Skorokhod, A.: Theory of Stochastic Processes. Springer, New York (2004–2007)
7. Horn, R., Johnson, C.: Matrix Analysis. Cambridge University Press, New York (1985)
8. Kato, T.: Perturbation Theory for Linear Operators. Springer, New York (1966)
9. Molchanov, S., Yarovaya, E.: The population structure inside the propagation front of a branching random walk with a finite number of particle generation centers. Dokl. Math. **86**(3), 787–790 (2012)
10. Vatutin, V., Topchiĭ, V.: Limit theorem for critical catalytic branching random walks. Theory Probab. Appl. **49**(3), 498–518 (2005)
11. Vatutin, V., Topchiĭ, V., Yarovaya, E.: Catalytic branching random walks and queueing systems with a random number of independent servers. Theory Probab. Math. Stat. **69**, 1–15 (2004)
12. Yarovaya, E.: Branching Random Walks in an Inhomogeneous Medium. Center of Applied Investigations of the Faculty of Mechanics and Mathematics of the Moscow State University, Moscow (2007) [in Russian]
13. Yarovaya, E.: About the investigation of branching random walks on multidimensional lattices. Sovrem. Probl. Mat. Mech. Mosc. **4**, 119–136 (2009) [in Russian]
14. Yarovaya, E.: Criteria of exponential growth for the numbers of particles in models of branching random walks. Theory Probab. Appl. **55**(4), 661–682 (2011)
15. Yarovaya, E.: Spectral properties of evolutionary operators in branching random walk models. Math. Notes **92**(1), 124–140 (2012)
16. Yarovaya, E.: Branching random walks with several sources. Math. Popul. Stud. **20**(1), 14–26 (2013)

Chapter 4
Branching Structures Within Random Walks and Their Applications

Wenming Hong and Huaming Wang

Mathematics Subject Classification (2000): 60J80, 60G50

4.1 Introduction

In 1975, Kesten, Kozlov and Spitzer published the famous paper [18] in which they proved a nice limit theorem for the nearest neighbor Random Walk in Random Environment (RWRE hereafter). They observed that the nearest neighbor RWRE, upon being properly centered and scaled, converges to the domain of attraction of certain stable distribution. The theorem that they proved will be called *stable limit theorem* throughout the paper. Its proof was based on the fact that there exists some branching structure within the path of the random walk. Roughly speaking, for the walk transient to the right, $1 = U_n, U_{n-1}, U_{n-2}, \ldots$ form a single-type branching process with exactly one immigrant in each of the first n generations, where for $k \le n-1$, U_k counts the number of steps by the walk from k to $k-1$ before it hits n. Then, by studying the limit behaviours of the single-type branching process with immigration in random environment, the limit theorem of RWRE could be studied. This approach was proved to be very useful for studying the nearest neighbor random walk in random or non-random environment. We provide here

---

W. Hong
School of Mathematical Sciences & Key Laboratory of Mathematics and Complex Systems,
Beijing Normal University, Beijing 100875, China
e-mail: wmhong@bnu.edu.cn

H. Wang (✉)
Department of Statistics, Anhui Normal University, Wuhu 241000, China
e-mail: hmking@mail.ahnu.edu.cn

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture
Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_4

some references. In Rogers [21] it is shown that, being properly scaled, the local time of the nearest neighbor random walk converges to Brownian local time. In [8], Gantert and Shi observed that, in transient case, the nearest neighbor RWRE spends, with a positive probability, a linear proportion of time at a single site. In [25], Zeitouni found the explicit density of the invariant equivalent measure related to the so-called "environment viewed from particles". Afanasyev strengthened in [1] Ritter's theorem to a functional setting and in the paper [2] he also proved a conditional Ritter's theorem. Note that all the results mentioned above highly rely on the branching structure hidden in the path of the nearest neighbor random walk.

It is attractive to generalize these results to some more general settings, for example, to random walks with bounded jumps. In this paper, the notation *"(L,R) random walk"* always denotes random walk with bounded jumps which in each step jumps at most a distance L to the left and at most a distance R to the right. The notation *"(L,R) RWRE"* could be understood similarly. In [19], Key studied the regeneration time for multitype branching process with immigration in random environment. Key indicated that the general argument for finding limiting distribution for $X(t)$ (He meant to generalize the stable limit theorem in [18] to (L,1) RWRE.) seems to go through, except now $Z(t)$ is multitype. Actually, it was true. One could generalize the stable limit theorem in [18] to (L,1) RWRE by studying a multitype branching process with immigration in random environment. However, the first obstacle that one has to overcome is to reveal the multitype branching process hidden in the path of (L,1) RWRE. For random walk with bounded jumps, things are very different from the nearest neighbor setting. Indeed, in [19], the author did not give the construction of the branching structure for (L,1) random walk.

In [11, 12, 14], we revealed the multitype branching process with immigration which is hidden in the path of random walk with bounded jumps, so that we could set up the connection between these two stochastic processes. The branching structure for (L,R) random walk is also proved to be powerful in applications of studying the limit behaviors for (L,R) random walk in random or non-random environment. We refer the readers to [10, 13, 16, 17, 22, 24] for the details of these applications.

The remainder parts of the paper are organized as follows. We devote Sect. 4.2 to a survey of the literatures related to the branching structure for (1,1) random walk. Then in Sect. 4.3, we list our results about the branching structures for random walks with bounded jumps set up in [11, 12]. After that, a birth-and-death process with bounded jumps will be considered in Sect. 4.4. Using the branching structure, we calculate the stationary distribution of the birth-and-death process explicitly. Finally, some concluding remarks will be given in Sect. 4.5.

4.2 Branching Structure for (1,1) Random Walk

Let us consider firstly the nearest neighbor random walk on $\mathbb{Z}$. Suppose that $p_k,\ k \in \mathbb{Z}$, is a sequence of [0, 1]-valued numbers. Let $\{S_n\}_{n\geq 0}$ be a Markov chain with $S_0 = 0$ and

$$P(S_{n+1} = k+1|S_n = k) = p_k = 1 - P(S_{n+1} = k-1|S_n = k).$$

We call $\{S_n\}$ the "*(1,1) random walk*". In the literatures that we are aware of, the history of the branching structure for (1,1) random walk could date back to Harris [9]. Harris found that there are some correspondences between random walks and trees. Putting $p_0 = 1$, he studied random walk on the lattice of the positive half-line. He illustrated that an excursion of the walk between time 0 and the first return time of the initial state corresponds to a tree. In [18], Kesten, Kozlov and Spitzer set up the immigration structures for (1,1) RWRE. They revealed the single-type branching process with immigration in random environment which is hidden in the path of (1,1) RWRE. Then, by applying the branching structure, they proved a stable limit theorem for (1,1) RWRE. In [7] the branching structure for (1,1) random walk with homogeneous transition probabilities was studied, that is, for all k, $p_k \equiv p \in (0, 1)$. For $a \geq 0$ let

$$N(a) := \#\{\text{steps by the walk from } a+1 \text{ to } a\}.$$

Here and throughout, the notation "#{ }" denotes the number of elements in set "{ }". Dwass proved that if $0 < p < \frac{1}{2}$, $N(0), N(1), \ldots$ form a Galton–Watson process; if $\frac{1}{2} < p < 1$, $N(0), N(1), \ldots$ evolve as a Galton–Watson process with certain immigration. We refer the readers to [7] for more details.

Next we quote the branching structure set up in [18] which influences us a lot. Suppose that $\limsup_{n\to\infty} S_n = \infty$. For $n \geq 1$, define

$$T_n = \inf\{k : S_k = n\},$$

which is the first hitting time of site n of the walk. Clearly $T_n < \infty$ for all $n > 0$. Let $U_n^n = 0$ and for $k \leq n-1$ set

$$U_k^n := \#\{\text{ steps by the walk from } k \text{ to } k-1 \text{ before } T_n\}.$$

The connection between the branching process with immigration and the (1,1) random walk in [18] reads as follows:

> Suppose that $\limsup_{n\to\infty} S_n = \infty$. Then $U_n^n, U_{n-1}^n, U_{n-2}^n, \ldots$ form a single type branching process with exactly 1 immigrant in each of the first n generations and no immigration in the remaining generations. The offspring distribution at $k+1$ is
>
> $$P(U_k^n = m|U_{k+1}^n = 1) = (1-p_k)^m p_k, m \geq 0.$$

Furthermore,

$$T_n = n + 2\sum_{i\le n} U_i^n. \tag{4.1}$$

Note that in (4.1), the hitting time T_n is written as the functional of the branching process $\{U_k^n\}_{k\le n}$. In [18], the authors considered the (1,1) RWRE. In this case, $\{U_k^n\}_{k\le n}$ forms a branching process with immigration in random environment. The stable limit theorem of the (1,1) RWRE was proved by studying the limit behaviour of $\{U_k^n\}_{k\le n}$. In [18] the authors explained why the above claims hold. The idea is very clear, but it is tedious to present a rigorous mathematical proof. We refer the readers to [11] for a rigorous proof in a more general setting.

4.3 Branching Structure for Random Walk with Bounded Jumps

In this section we introduce the branching structure within the path of random walk with bounded jumps. To define the model, let $L, R \ge 1$ and denote $\Lambda = \{-L, \ldots, R\} \setminus \{0\}$. For $i \in \mathbb{Z}$, let $\omega_i = (\omega_i(l))_{l\in\Lambda}$ be a probability measure on $i + \Lambda$. Set $\omega = \{\omega_i\}_{i\in\mathbb{Z}}$, which serves as the transition probabilities of the random walk. Let $\{X_n\}_{n\ge 0}$ be a Markov chain with initial value $X_0 = x$ and transition probabilities

$$P_\omega(X_{n+1} = i + j \big| X_n = i) = \omega_i(j),\ j \in \Lambda.$$

$\{X_n\}$ is called the *(L,R) random walk with non-homogeneous transition probabilities* ((L,R) random walk). Throughout this section, we use P_ω^x to denote the probability distribution induced by the random walk $\{X_n\}$ and E_ω^x the corresponding integration operator. Except otherwise stated, we always assume that the walk $\{X_n\}$ starts at 0. The notations P_ω^0 and E_ω^0 will be written as P_ω and E_ω, respectively, for simplicity.

In [12], we set up the branching structure for (2,2) random walk. For a general (L,R) random walk, the idea is the same, but the notations are much heavier. The branching structure for (L,1) random walk and (1,R) random walk were set up in [11] and [14], respectively. Here we emphasize that the branching structure for (L,1) random walk is almost as nice as that of (1,1) random walk in the sense that the offspring distributions of the individuals at i depend only on the transitional probabilities of the random walk at i (See (4.2) and (4.3) below). Therefore many related results of (1,1) random walk in random or non-random environment could be possibly generalized to the (L,1) setting.

4.3.1 (L,1) Random Walk

Letting $R = 1$, we consider (L,1) random walk $\{X_n\}$. For $n \geq 0$, define the hitting time of site n by the walk by

$$T_n = \inf[k \geq 0 : X_k = n].$$

In general, it is hard to find the exact distribution of T_n. In Theorem 4.1 we will see that T_n could be expressed in terms of a multitype branching process with immigration in varying environment.

For $-\infty < i < n,\ 1 \leq l \leq L$, define

$$U_{i,l}^n = \#\{0 < k < T_n : X_{k-1} > i, X_k = i - l + 1\},$$

and set

$$U_i^n := (U_{i,1}^n, U_{i,2}^n, \cdots, U_{i,L}^n).$$

Throughout the paper, for $i \geq 1,\ \mathbf{e}_i$ denotes the row vector whose ith component equals 1 and all other components equal 0.

Theorem 4.1 *Suppose that* $\limsup_{n\to\infty} X_n = \infty$. *Then for* $n > 0$*,*

(1) $T_n = n + \sum_{i=-\infty}^{n-1} U_i^n \mathbf{x}_0$, *where* $\mathbf{x}_0 = (2, 1, \ldots, 1)^t$;
(2) the process $\{U_i^n\}_{i<n}$ *evolves as an L-type branching process with immigration. The immigration satisfies the following properties: with probability* 1*, a type-*1 *particle immigrates at k for* $0 \leq k \leq n-1$, *and there is no immigration at k for* $k \geq n$ *and* $k < 0$. *Furthermore, the offspring distributions of the multitype branching process* $\{U_i^n\}_{i<n}$ *are as follows: for* $(u_1, \ldots, u_L) \in \mathbb{Z}_+^L$,

$$P_\omega(U_{i-1}^n = (u_1, \ldots, u_L) \big| U_i^n = \mathbf{e}_1) = \frac{(u_1 + \ldots + u_L)!}{u_1! \cdots u_L!} \omega_i(-1)^{u_1} \cdots \omega_i(-L)^{u_L} \omega_i(1), \tag{4.2}$$

$$P_\omega\left(U_{i-1}^n = \mathbf{e}_{l-1} + (u_1, \ldots, u_L) \big| U_i^n = \mathbf{e}_l\right) = \frac{(u_1 + \ldots + u_L)!}{u_1! \cdots u_L!} \omega_i(-1)^{u_1} \cdots \omega_i(-L)^{u_L} \omega_i(1),\ 2 \leq l \leq L. \tag{4.3}$$

We see from (4.2) and (4.3) that the offspring distributions of individuals at i only depend on ω_i. This fact is very important for applications. Indeed, if $\{\omega_i\}$ is an i.i.d. sequence under some probability measure $\mathbb{P}$, then $\{U_i\}_{i\leq n}$ forms an L-type branching process with immigration in i.i.d. random environment. Consequently, we could use directly the limit theorems of the later one to study the limit behaviours of (L,1) RWRE. In [22], we studied the tail behaviours of the total progenies of a multitype branching process with immigration in i.i.d. random environment before the regenerating time, and proved a stable limit theorem for (L,1) RWRE, which generalized the stable limit theorem for (1,1) RWRE derived in [18]. Here we also introduce some other applications. In [10], Hong and Sun proved a renewal theorem for (L,1) RWRE. In [16, 17] the stationary distributions of (L,1) random walk and (1,R) random walk were studied respectively. Hong and Yang [13] showed that the local time of (1,L) random walk converges to Brownian local time. All these applications highly rely on the branching structure for (L,1) random walk. In Sect. 4.4, we will also discuss the stationary distribution for a birth-and-death process with one-side bounded jumps in details to illustrate further application of the branching structure for (L,1) random walk.

4.3.2 (2,2) Random Walk

Fixing $L = R = 2$, we consider the (2,2) random walk. Define

$$T_1 = \inf[n \geq 0 : X_n > 0],$$

which is of special interest in the study of random walk with bounded jumps in random environment.

In order to count exactly all the steps of the walk before T_1, we define three types of excursions.

Definition 4.1

- **a)** We call excursions of the form $\{X_k = i, X_{k+1} = i-1, X_{k+2} \leq i-1, \ldots, X_{k+l} \leq i-1, X_{k+l+1} \geq i\}$ type-$\mathcal{A}$ excursions at i. We classify type-$\mathcal{A}$ excursions at i into three sub-types $\mathcal{A}_{i,1}$, $\mathcal{A}_{i,2}$ and $\mathcal{A}_{i,3}$, corresponding to the possible three kinds of last step of type-$\mathcal{A}$ excursions at i, say, $\{i-1 \to i\}$, $\{i-2 \to i\}$ and $\{i-1 \to i+1\}$.
- **b)** We call excursions of the form $\{X_k = i, X_{k+1} = i-2, X_{k+2} \leq i-1, \ldots, X_{k+l} \leq i-1, X_{k+l+1} \geq i\}$ type-$\mathcal{B}$ excursions at i. We classify type-$\mathcal{B}$ excursions at i into three sub-types $\mathcal{B}_{i,1}$, $\mathcal{B}_{i,2}$ and $\mathcal{B}_{i,3}$, corresponding to the possible three kinds of last step of type-$\mathcal{B}$ excursions at i, say, $\{i-1 \to i\}$, $\{i-2 \to i\}$ and $\{i-1 \to i+1\}$.
- **c)** We call excursions of the form $\{X_k = i+1, X_{k+1} = i-1, X_{k+2} \leq i-1, \ldots, X_{k+l} \leq i-1, X_{k+l+1} \geq i\}$ type-C excursions at i. We classify type-C excursions at i into three sub-types $C_{i,1}$, $C_{i,2}$ and $C_{i,3}$, corresponding to the possible three kinds of last step of type-C excursions at i, say, $\{i-1 \to i\}$, $\{i-2 \to i\}$ and $\{i-1 \to i+1\}$ (Fig. 4.1).

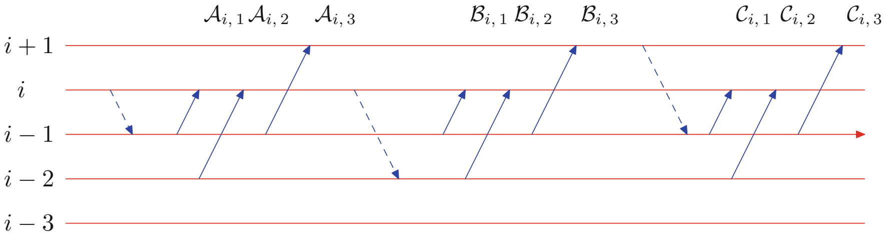


Fig. 4.1 The figure illustrates type $\mathcal{A}$, $\mathcal{B}$ and C excursions at i. We only draw the first step and the last step. All other steps below $i-1$ are omitted

For $i \le 0$ and $j = 1, 2, 3$, define

$$\begin{aligned} A_{i,j} &= \#\{\mathcal{A}_{i,j} \text{ excursions before } T_1\}, \\ B_{i,j} &= \#\{\mathcal{B}_{i,j} \text{ excursions before } T_1\}, \\ C_{i,j} &= \#\{C_{i,j} \text{ excursions before } T_1\}. \end{aligned}$$

To count all the steps of the walk before T_1, let

$$U_i = (A_{i,1}, A_{i,2}, A_{i,3}, B_{i,1}, B_{i,2}, B_{i,3}, C_{i,1}, C_{i,2}, C_{i,3}) \tag{4.4}$$

denote the total numbers of different excursions at i before time T_1. Then, we have the following fact.

Theorem 4.2 *Suppose that* $\limsup_{n\to\infty} X_n = \infty$. *Then*

$$T_1 = 1 + \sum_{i\le 0} U_i(2, 2, 1, 1, 1, 0, 2, 2, 1)^t. \tag{4.5}$$

Remark 4.1 Since the (2,2) random walk is non-nearest neighbor, one can not provide the exact distribution of T_1 in general; but the process $\{U_i\}_{i\le 0}$ defined in (4.4) will be proved to be a non-homogeneous multitype branching process. This fact together with (4.5) enables us to study T_1 by the properties of the branching process.

In order to prove that $\{U_i\}_{i\le 0}$ forms a multitype branching process, we need to formulate an ancestor. Note that the walk starts from 0, but before T_1, there is no step downward from some site above 0 to 0. Therefore we can assume that there is a step of the walk from 1 to 0 before it starts from 0 (We can also assume that this step is from 2 to 0, but it makes no difference.), that is, set $X_{-1} = 1$. Adding such an imaginary step, the path $\{X_{-1} = 1, X_0 = 0, X_1, \dots, X_{T_1}\}$ forms a type-$\mathcal{A}$ excursion at 1 such that with probability 1,

$$A_{1,1} + A_{1,2} + A_{1,3} = 1.$$

Then U_1 could be defined as in (4.4). But since there is no $\mathscr{B}_{1,j}$ and $C_{1,j}$, $j = 1, 2, 3$, excursion, U_1 has only three possible values, that is, $U_1 = \mathbf{e}_1$, $U_1 = \mathbf{e}_2$ or $U_1 = \mathbf{e}_3$.

We can treat U_1 as some particle which immigrates into the system and call it *immigration*. To reveal the immigration law, for $k \le i$ and $j \in \{i+1, i+2\}$, we define the escaping probabilities

$$f_k(i,j) = P_\omega^k(\text{the walk hits } (i,\infty) \text{ at } j),$$

which could be explicitly calculated by some standard procedure. In [12], we showed that

$$P_\omega^0(U_1 = \mathbf{e}_1) = \frac{\omega_0(1)}{1 - \omega_0(-1)f_{-1}(-1,0) - \omega_0(-2)f_{-2}(-1,0)}, \tag{4.6}$$

$$P_\omega^0(U_1 = \mathbf{e}_2) = \frac{\omega_0(2)}{1 - \omega_0(-1)f_{-1}(-1,0) - \omega_0(-2)f_{-2}(-1,0)}, \tag{4.7}$$

$$P_\omega^0(U_1 = \mathbf{e}_3) = \frac{1 - \omega_0(1) - \omega_0(2) - \omega_0(-1)f_{-1}(-1,0) - \omega_0(-2)f_{-2}(-1,0)}{1 - \omega_0(-1)f_{-1}(-1,0) - \omega_0(-2)f_{-2}(-1,0)}. \tag{4.8}$$

Theorem 4.3 *Suppose that* $\limsup_{n\to\infty} X_n = \infty$. *Then* $\{U_i\}_{i\le 1}$ *forms a 9-type non-homogeneous branching process whose immigration distributions are given by (4.6), (4.7) and (4.8).*

We did not present the exact offspring distributions of the multitype branching process in Theorem 4.3 because they are very complicated. To formulate the offspring distributions, we need to define an index for each one of the nine types of excursions. These indices are functionals of the escaping probabilities. With the help of these indices, the offspring distribution of each one of the nine types of excursions could be formulated. We refer the readers to [12] for the proofs of Theorems 4.2 and 4.3.

The branching structure for (L,R) random walk is also proved to be useful. We provide here some examples. We also look forward for some other applications.

(i) *To formulate the explicit velocity for the law of large numbers of the random walk with bounded jumps in random environment.* For (L,1) RWRE, it's always possible to find the explicit velocity for the law of large numbers, see Brémont [4]. However, for (L,R) RWRE, though Brémont also proved a law of large numbers in [5], no explicit velocity is available any longer. In [12], by using the branching structure for (L,R) random walk, we could formulate the explicit invariant density for the measure of the so-called "environment viewed from particles", which is equivalent to the original measure of the environment.

Consequently, we could prove the law of large number for (L,R) RWRE and formulate its velocity explicitly. Similar method also applies to birth-and-death process with bounded jumps in random environment. For details, see [24].

(ii) *To provide the explicit stationary distribution for the random walk with nonhomogeneous transition probabilities on the lattice of the positive half-line.* The idea is as follows. Consider an excursion of the walk between time 0 and the first return time of site 0. Using the branching structure, the first return time of site 0 and the occupation time at some location i of the walk could be written as the functionals of a multitype branching process, consequently one could calculate their means. The stationary distribution could be formulated taking into account that the proportion of the mean occupation time at i with respect to the mean of the first return time of site 0 is just the mass of the stationary distribution at i.

4.4 Stationary Distribution for a Birth-and-Death Process with One-Side Bounded Jumps

In this section, to illustrate the application of the branching structure for (R,1) random walk introduced in Sect. 4.3.1, we formulate the stationary distribution for a birth-and-death process with one-side bounded jumps. To introduce the model, fix $1 \le R \in \mathbb{Z}$ and let $(\mu_i, \lambda_i^1, \ldots, \lambda_i^R)_{i\ge 0}$ be a sequence of nonnegative $\mathbb{R}^{R+1}$-valued vectors, where $\mu_0 = 0$. Let $\{N_t\}_{t\ge 0}$ be a continuous time Markov chain, with waiting time at the state n exponentially distributed with parameter $\mu_n + \sum_{r=1}^R \lambda_n^r$. Once it leaves state n, it jumps to $n + i$ with probability $\lambda_n^i/(\mu_n + \sum_{r=1}^R \lambda_n^r),\ i = 1, \ldots, R$ or to $n - 1$ with probability $\mu_n/(\mu_n + \sum_{r=1}^R \lambda_n^r)$. The process $\{N_t\}_{t\ge 0}$ is called a *nonhomogeneous birth-and-death process with one-side bounded jumps*.

The process $\{N_t\}$ defined above is the continuous time analogue of the so-called (1,R) random walk which was studied in Hong and Zhou [14]. The stationary distribution of the walk is also provided in the aforementioned paper.

First we present some sufficient conditions for recurrence and positive recurrence. Then, for the positive recurrent case, we formulate the explicit stationary distribution of $\{N_t\}$. The idea to formulate the stationary distribution is to look only at the discontinuities of $\{N_t\}$ in order to get its embedded process $\{X_n\}$, whose stationary distribution $\pi_k, k \ge 0$ could be formulated by mean of the branching structure introduced in Sect. 4.3.1. Since $\{\pi_k/(\mu_k + \sum_{r=1}^R \lambda_k^r)\}_{k\ge 0}$ defines an invariant measure for the process $\{N_t\}$, the stationary distribution of the process could be formulated. The following condition is needed.

(C) There are $\kappa > 0$ and $K < \infty$ such that for all $n > 0,\ \kappa < \mu_n < K$ and for $n \ge 0,\ \kappa < \sum_{r=1}^R \lambda_n^r < K$.

Remark 4.2 Let $Q = (q_{ij})_{i,j\geq 0}$ be a matrix with

$$q_{ij} = \begin{cases} \mu_i, & \text{if } j = i-1; \\ \lambda_i^r, & \text{if } j = i+r,\ r = 1, \ldots, R; \\ -\left(\mu_i + \sum_{r=1}^R \lambda_i^r\right), & \text{if } j = i; \\ 0, & \text{else.} \end{cases}$$

Then, from condition (C), it follows that the matrix Q is a conservative Q-matrix bounded from above. Hence the process $\{N_t\}$ exists (see Anderson [3, Proposition 2.9, Chap. 2].). Some weaker condition can also imply its existence (see Wang [23]).

For each $i \geq 1$, define $a_i^k = \frac{\sum_{l=k}^R \lambda_i^l}{\mu_i}$, $k = 1, \ldots, R$,

$$M_i = \begin{pmatrix} a_i^1 & \cdots & a_i^{R-1} & a_i^R \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix},$$

and

$$P_t(i,j) = P(N_t = j | N_0 = i).$$

The following theorem is the main result of this section.

Theorem 4.4 *Under condition (C), the following statements hold.*

(a) If $\lim_{n\to\infty} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1} \mathbf{e}_1^t = 0$*, then* $\{N_t\}$ *is recurrent.*
(b) If $\sum_{n=1}^\infty \frac{1}{\mu_n} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1} \mathbf{e}_1^t < \infty$*, then* $\{N_t\}$ *is positive recurrent and the limits*

$$\psi_0 := \lim_{t\to\infty} P_t(i,0) = \frac{(\sum_{r=1}^R \lambda_0^r)^{-1}}{(\sum_{r=1}^R \lambda_0^r)^{-1} + \sum_{n=1}^\infty \frac{1}{\mu_n} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1} \mathbf{e}_1^t}, \tag{4.9}$$

$$\psi_k := \lim_{t\to\infty} P_t(i,k) = \frac{\frac{1}{\mu_k} \mathbf{e}_1 M_1 M_2 \cdots M_{k-1} \mathbf{e}_1^t}{(\sum_{r=1}^R \lambda_0^r)^{-1} + \sum_{n=1}^\infty \frac{1}{\mu_n} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1} \mathbf{e}_1^t}, \tag{4.10}$$

define a stationary distribution for $\{N_t\}$ *in the sense that for all* $t > 0$*,*

$$\psi_k = \sum_{n=0}^\infty \psi_n P_t(n,k),\ k \geq 0.$$

Remark 4.3 Previously to the proof of the theorem, recall the meaning of recurrence and positive recurrence of a continuous time Markov chain $\{N_t\}$. Let $\tau = \inf\{t > 0 : N_t \neq N_0\}$ and define $\eta = \inf\{t > \tau : N_t = N_0\}$. By definition, τ is the time when $\{N_t\}$ leaves the starting state and η is the time when $\{N_t\}$ returns to the starting state for the first time after having left it. If $P(\eta < \infty) = 1$, we say that $\{N_t\}$ is recurrent; if $E\eta < \infty$, we say that $\{N_t\}$ is positive recurrent.

Proof Note that under condition (C), $\{N_t\}$ exists. Let $\tau_0 = 0$ and define recursively for $n \geq 1$,

$$\tau_n = \inf\{t \geq \tau_{n-1} : N_t \neq N_{\tau_{n-1}}\},$$

with the convention that $\inf \phi = \infty$. For $n \geq 0$, set $X_n = N_{\tau_n}$. Then, $\{X_n\}_{n\geq 0}$ is a discrete time Markov chain on the lattice of the positive half-line with transition probabilities

$$r_{ij} = \begin{cases} \frac{\mu_i}{\mu_i + \sum_{l=1}^R \lambda_i^l}, & j = i-1 \\ \frac{\lambda_i^k}{\mu_i + \sum_{l=1}^R \lambda_i^l}, & j = i+k, k = 1, .., R \\ 0, & \text{else.} \end{cases}$$

The process $\{X_n\}$ is also known as the *embedded process* of $\{N_t\}$.

Denote by

$$T = \inf\{k > 0 : X_k = X_0\}$$

the time when $\{X_n\}$ returns to X_0 for the first time, and by

$$\eta = \inf\{t > \tau_1 : N_t = N_0\}$$

the time when $\{N_t\}$ returns to N_0 for the first time after having left N_0.

To prove part (a) of the theorem, it suffices to show that $\{X_n\}$ is recurrent whenever $\lim_{n\to\infty} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1} \mathbf{e}_1^t = 0$. For this purpose, define for $0 \leq a < k < b$,

$$P_k(a, b, +) := P(\text{starting from } k, \{X_n\} \text{ exits } [a+1, b-1] \text{ from above}).$$

From Markov properties it follows that

$$P_k(a, b, +) = \frac{\sum_{j=a+1}^k \mathbf{e}_1 M_j M_{j+1} \cdots M_{b-1} \mathbf{e}_1^t}{\sum_{j=a+1}^b \mathbf{e}_1 M_j M_{j+1} \cdots M_{b-1} \mathbf{e}_1^t},$$

where the empty product is the identity. Thus we have that for $b > k$,

$$P_k(\{X_n\} \text{ hits } k-1 \text{ before it hits } [b,\infty)) = 1 - \frac{\mathbf{e}_1 M_k M_{k+1} \cdots M_{b-1}\mathbf{e}_1^t}{1 + \sum_{j=k}^{b-1} \mathbf{e}_1 M_j M_{j+1} \cdots M_{b-1}\mathbf{e}_1^t}.$$

Since $\lim_{n\to\infty} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1}\mathbf{e}_1^t = 0$, from the condition (C), it follows that

$$P_k(\{X_n\} \text{ hits } k-1 \text{ for some } n \geq 0) = 1.$$

Thus, $P(T < \infty) = 1$. Consequently, $\{X_n\}$ is recurrent and so is $\{N_t\}$.

To prove part (b) of the theorem, suppose that $\sum_{n=1}^{\infty} \frac{1}{\mu_n}\mathbf{e}_1 M_1 M_2 \cdots M_{n-1}\mathbf{e}_1^t < \infty$. Since $\mu_n^{-1}, n \geq 1$ are uniformly bounded away from zero, we have that

$$\lim_{n\to\infty} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1}\mathbf{e}_1^t = 0.$$

Consequently, it follows from part (a) of the theorem that both $\{X_n\}$ and $\{N_t\}$ are recurrent. Therefore, $P(\eta < \infty) = P(T < \infty) = 1$.

To show that $\{N_t\}$ is positive recurrent, it is sufficient to prove $E(\eta) < \infty$. Since $\{N_t\}$ is recurrent, without loss of generality, we can assume that $N_0 = 0$. Let $U_1 = \mathbf{e}_1$, and for each $i \geq 2$, define

$$U_{i,r} = \#\{0 < k < T : X_{k-1} < i, X_k = i + r - 1\},\ r = 1, \ldots, R,$$

and let

$$U_i := (U_{i,1}, U_{i,2}, \cdots, U_{i,R}).$$

It follows from Theorem 4.1 in Sect. 4.3.1 that $(U_n)_{n\geq 1}$ forms an R-type branching process with offspring distributions

$$P(U_{i+1} = (u_1, \ldots, u_R) \big| U_i = \mathbf{e}_1)$$
$$= \frac{(u_1 + \ldots + u_R)!}{u_1! \cdots u_R!} \left(\frac{\lambda_i^1}{\mu_i + \sum_{k=1}^R \lambda_i^k}\right)^{u_1} \cdots \left(\frac{\lambda_i^R}{\mu_i + \sum_{k=1}^R \lambda_i^k}\right)^{u_R} \left(\frac{\mu_i}{\mu_i + \sum_{k=1}^R \lambda_i^k}\right), \tag{4.11}$$

and for $2 \leq l \leq R$,

$$P\left(U_{i+1} = \mathbf{e}_{l-1} + (u_1, \ldots, u_R) \big| U_i = \mathbf{e}_l\right)$$
$$= \frac{(u_1 + \ldots + u_R)!}{u_1! \cdots u_R!} \left(\frac{\lambda_i^1}{\mu_i + \sum_{k=1}^R \lambda_i^k}\right)^{u_1} \cdots \left(\frac{\lambda_i^R}{\mu_i + \sum_{k=1}^R \lambda_i^k}\right)^{u_R} \left(\frac{\mu_i}{\mu_i + \sum_{k=1}^R \lambda_i^k}\right). \tag{4.12}$$

By some careful calculation, from (4.11) and (4.12), one has that

$$E(U_i) = \mathbf{e}_1 A_1 A_2 \cdots A_{i-1}, \tag{4.13}$$

where

$$A_i = \begin{pmatrix} b_i^1 & \ldots & b_i^{R-1} & b_i^R \\ 1 + b_i^1 & \ldots & b_i^{R-1} & b_i^R \\ \vdots & \ddots & \vdots & \vdots \\ b_i^1 & \ldots & 1 + b_i^{R-1} & b_i^R \end{pmatrix},$$

with $b_i^r = \frac{\lambda_i^r}{\mu_i}$, $r = 1, \ldots, R$.

Considering the occupation time of $\{X_n\}$ before T, one has that $\sum_{k=0}^{T-1} 1_{X_k=0} = 1$ and for $i \geq 1$,

$$\sum_{k=0}^{T-1} 1_{X_k=i} = U_{i,1} + U_{i+1}\mathbf{1}, \tag{4.14}$$

where here and throughout $\mathbf{1} := (\mathbf{e}_1 + \ldots + \mathbf{e}_R)^t$.

Considering the occupation time of $\{N_t\}$ before η, one has $\int_0^\eta 1_{N_t=0} dt = \xi_{0,1}$ and for $i \geq 1$,

$$\int_0^\eta 1_{N_t=i} dt = \sum_{k=1}^{U_{i,1}+U_{i+1}\mathbf{1}} \xi_{i,k}, \tag{4.15}$$

where $\xi_{i,k}, i \geq 0, k \geq 1$ are mutually independent random variables, which are also independent of U_i, and such that $P(\xi_{i,k} > t) = e^{-t(\mu_i + \sum_{r=1}^R \lambda_i^r)}, t \geq 0$. For the proof of (4.14) and (4.15), we refer the reader to Wang [23].

By Ward's equation (see Theorem 4.1.5, in Durrett [6]), from (4.13) and (4.15), it follows that

$$\begin{aligned} E(\eta) &= \sum_{n=0}^\infty E\Big(\int_0^\eta 1_{N_t=n} dt\Big) = E\xi_{0,1} + \sum_{n=1}^\infty E(U_{n,1} + U_{n+1}\mathbf{1}) E\xi_{n,1} \\ &= \Big(\sum_{r=1}^R \lambda_0^r\Big)^{-1} + \sum_{n=1}^\infty \frac{1}{\mu_n + \sum_{r=1}^R \lambda_n^r} (\mathbf{e}_1 A_1 A_2 \cdots A_{n-1} \mathbf{e}_1^t + \mathbf{e}_1 A_1 A_2 \cdots A_n \mathbf{1}) \\ &= \Big(\sum_{r=1}^R \lambda_0^r\Big)^{-1} + \sum_{n=1}^\infty \frac{1}{\mu_n} \mathbf{e}_1 A_1 A_2 \cdots A_{n-1} \mathbf{1} \\ &= \Big(\sum_{r=1}^R \lambda_0^r\Big)^{-1} + \sum_{n=1}^\infty \frac{1}{\mu_n} \mathbf{e}_1 M_1 M_2 \cdots M_{n-1} \mathbf{e}_1^t < \infty. \end{aligned}$$

Therefore, $\{N_t\}$ is positive recurrent.

The existence of the limits ψ_k in (4.9) and (4.10) follows from Theorem 1.6 of Chap. 5 in Anderson [3]. In the same theorem, it is also proved that if $(\psi_k)_{k\geq 0}$ is a probability distribution, then it is the unique probability distribution such that for all $t > 0$,

$$\psi_k = \sum_{n=0}^{\infty} \psi_n P_t(n,k),\ k \geq 0. \tag{4.16}$$

On the other hand, in Theorem 3.5.1 of Norris [20], it is shown that $\left(\frac{\pi_k}{\mu_k+\sum_{r=1}^R \lambda_k^r}\right)_{k\geq 0}$ satisfies (4.16), where $(\pi_k)_{k\geq 0}$ is the stationary distribution of the embedded process $\{X_n\}$. Next we calculate $(\pi_k)_{k\geq 0}$. For the (1,R) random walk which could stay at its site, $(\pi_k)_{k\geq 0}$ was studied in Hong and Zhou [14]. Since the calculation is not too long and the notations here are a bit different from those in [14], we repeat the calculation of $(\pi_k)_{k\geq 0}$.

Note that by (4.13) and (4.14), $E(\sum_{k=0}^{T-1} 1_{X_k=0}) = 1$ and for $n \geq 1$,

$$\begin{aligned}
E\Big(\sum_{k=0}^{T-1} 1_{X_k=n}\Big) &= E(U_{n,1} + U_{n+1}\mathbf{1}) \\
&= \mathbf{e}_1(A_1A_2\cdots A_{n-1}\mathbf{e}_1^t + A_1A_2\cdots A_n\mathbf{1}) \\
&= \frac{\mu_n + \sum_{r=1}^R \lambda_n^r}{\mu_n}\mathbf{e}_1A_1A_2\cdots A_{n-1}\mathbf{1} \\
&= \frac{\mu_n + \sum_{r=1}^R \lambda_n^r}{\mu_n}\mathbf{e}_1M_1M_2\cdots M_{n-1}\mathbf{e}_1^t.
\end{aligned}$$

One also has

$$\begin{aligned}
ET &= 1 + \sum_{n=1}^{\infty} E(U_{n,1} + U_{n+1}\mathbf{1}) \\
&= 1 + \sum_{n=1}^{\infty} \frac{\mu_n + \sum_{r=1}^R \lambda_n^r}{\mu_n}\mathbf{e}_1M_1M_2\cdots M_{n-1}\mathbf{e}_1^t.
\end{aligned}$$

Let $\pi_0 = \frac{1}{1+\sum_{n=1}^{\infty}\frac{\mu_n+\sum_{r=1}^R\lambda_n^r}{\mu_n}\mathbf{e}_1M_1M_2\cdots M_{n-1}\mathbf{e}_1^t}$ and for $k \geq 1$ let

$$\pi_k = \frac{\frac{\mu_k+\sum_{r=1}^R\lambda_k^r}{\mu_k}\mathbf{e}_1M_1M_2\cdots M_{k-1}\mathbf{e}_1^t}{1+\sum_{n=1}^{\infty}\frac{\mu_n+\sum_{r=1}^R\lambda_n^r}{\mu_n}\mathbf{e}_1M_1M_2\cdots M_{n-1}\mathbf{e}_1^t}.$$

Then, one has that $(\pi_k)_{k\geq 0}$ defines a stationary distribution for $\{X_n\}$ (see Durrett [6]). Set $\nu_k = \pi_k/(\mu_k + \sum_{r=1}^R \lambda_k^r)$, for $k \geq 0$. Then, $(\nu_k)_{k\geq 0}$ satisfies (4.16). Normalizing ν_k by setting $\psi_k = \nu_k/(\sum_{k=0}^\infty \nu_k)$ one obtains (4.9) and (4.10). We conclude that $\psi_k, k \geq 0$, is the stationary distribution of $\{N_t\}$ and hence the theorem is proved. □

4.5 Concluding Remarks

(a) We do not present the branching structure for (1,R) random walk constructed in [14] in detail since it is a special case of (L,R) random walk. However, it is important to mention that our construction of the branching structure for (L,R) random walk was strongly influenced by [14]. Some more general model has been considered as well. In [15], Hong and Zhang established the branching structure for the nearest neighbor random walk on a strip.

(b) The branching structure for (L,1) random walk could be generalized to a class of random walk with unbounded jumps, which could jump a distance of any size downwards or a distance 1 upwards in each step. Precisely, for $i \in \mathbb{Z}$, let $(p_i, q_{i1}, q_{i2}, \ldots)$ be an infinite vector with $p_i, q_{ij} \geq 0, j \in \mathbb{Z}$ and $p_i + \sum_{j=1}^\infty q_{ij} = 1$, and consider random walk $\{X_n\}$ with $X_0 = 0$ and transition probabilities

$$P(X_{n+1} = j | X_n = i) = \begin{cases} p_i, & j = i+1 \\ q_{ik}, & j = i-k, k = 1, 2, \ldots. \end{cases}$$

Let $T = \inf\{n > 0 : X_n = 1\}$ and for $i < 0,\ k \geq 1$, define

$$U_{i,k} = \#\{0 < n < T : X_{n-1} > i, X_n = i - k + 1\}.$$

Let $U_0 = (1, 0, 0, \ldots)$ and for $i < 0$, denote

$$U_i = (U_{i,1}, U_{i,2}, U_{i,3}, \ldots).$$

Then, one has the following result:

> Suppose that $\max_{n\geq 0} X_n > 0$. Then $T = 1 + \sum_{i<0} U_i (2, 1, 1, \ldots)^t$. Furthermore $U_0, U_{-1}, U_{-2}, \ldots$ form a countable-type branching process whose offspring distributions are as follows: for $(u_1, u_2, u_3 \ldots) \in \mathbb{Z}_+^{\mathbb{N}}$ with $\sum_{k=1}^\infty u_i < \infty$,
>
> $$P(U_{i-1} = (u_1, u_2, u_3 \ldots) | U_i = \mathbf{e}_1) = \frac{(u_1 + u_2 + \ldots)!}{u_1! u_2! u_3! \cdots} p_i \prod_{j=1}^\infty q_{ij}^{u_j},$$
>
> $$P(U_{i-1} = \mathbf{e}_{k-1} + (u_1, u_2, u_3 \ldots) | U_i = \mathbf{e}_k) = \frac{(u_1 + u_2 + \ldots)!}{u_1! u_2! u_3! \cdots} p_i \prod_{j=1}^\infty q_{ij}^{u_j},\ k \geq 2.$$

The proof of this result is similar to that of Theorem 4.1 (we refer the reader to [11]).

Acknowledgements The authors are grateful to the referee for his or her careful reading and valuable suggestions. The project is supported in part by National Natural Science Foundation of China (Grant Nos. 11131003, 11501008), Nature Science Foundation of Anhui Province (Grant No. 1508085QA12) and Nature Science Foundation of Anhui Educational Committee (Grant No. KJ2014A085).

References

1. Afanasyev, V.I.: About time of reaching a high level by a random walk in a random environment. Theory Probab. Appl. **57**, 547–567 (2013)
2. Afanasyev, V.I.: Conditional limit theorem for the maximum of random walk in random environment. Theory Probab. Appl. **58**, 525–545 (2014)
3. Anderson, W.J.: Continuous-Time Markov Chains: An Applications-Oriented Approach. Springer, New York (1991)
4. Brémont, J.: On some random walks on $\mathbb{Z}$ in random medium. Ann. Probab. **30**, 1266–1312 (2002)
5. Brémont, J.: One-dimensional finite range random walk in random medium and invariant measure equation. Ann. Inst. Henri Poincaré Probab. Stat. **45**, 70–103 (2009)
6. Durrett, R.: Probability: Theory and Examples. 4th edn. Cambridge University Press, New York (2010)
7. Dwass, M.: Branching processes in simple random walk. Proc. Am. Math. Soc. **51**, 270–274 (1975)
8. Gantert, N., Shi, Z.: Many visits to a single site by a transient random walk in random environment. Stoch. Process. Appl. **99**, 159–176 (2002)
9. Harris, T.E.: First passage and recurrence distributions. Trans. Am. Math. Soc. **73**, 471–486 (1952)
10. Hong, W.M., Sun, H.Y.: Renewal theorem for (L;1)-random walk in random environment. Acta Math. Sci. **33**, 1736–1748 (2013)
11. Hong, W.M., Wang, H.M.: Intrinsic branching structure within (L-1) random walk in random environment and its applications. Infinite Dimens. Anal. Quantum Probab. Relat. Top. **16**, 1350006 [14 pages] (2013)
12. Hong, W.M., Wang, H.M.: Intrinsic branching structure within random walk on Z. Theory Probab. Appl. **58**, 640–659 (2014)
13. Hong, W.M., Yang, H.: Scaling limit of the local time of the (1,L)–random walk. ArXiv: 1402.3949 (2014)
14. Hong, W.M., Zhang, L.: Branching structure for the transient (1;R)–random walk in random environment and its applications. Infinite Dimens. Anal. Quantum Probab. Relat. Top. **13**, 589–618 (2010)
15. Hong, W.M., Zhang, M.J.: Branching structure for the transient random walk on a strip in a random environment. arXiv:1204.1104 (2014)
16. Hong, W.M., Zhou, K.: Tail asymptotic of the stationary distribution for the state dependent (1,R)-reflecting random walk-near critical. arXiv:1302.3069 (2013)
17. Hong, W.M., Zhou, K., Zhao, Y.Q.Q.: Explicit stationary distribution of the (L,1)-reflecting random walk on the half line. Acta Math. Sci. **30**, 371–388 (2014)
18. Kesten, H., Kozlov, M.V., Spitzer, F.: A limit law for random walk in a random environment. Compos. Math. **30**, 145–168 (1975)

19. Key, E.S.: Limiting distributions and regeneration times for multitype branching processes with immigration in a random environment. Ann. Probab. **15**, 344–353 (1987)
20. Norris, J.R.: Markov Chains. Cambridge University Press, Cambridge (1997)
21. Rogers, L.C.G.: Brownian local times and branching processes. In: Séminaire de Probabilités XVIII 1982/83. Lecture Notes in Mathematics, vol. 1059, pp. 42–55. Springer, Heidelberg (1984)
22. Wang, H.M.: A note on multitype branching process with bounded immigration in random environment. Acta Math. Sci. **29**, 1095–1110 (2013)
23. Wang, H.M.: Birth and death process with one-side bounded jumps in random environment. arXiv:1407.3385 (2014)
24. Wang, H.M.: Law of large numbers for random walk with unbounded jumps and BDP with bounded jumps in random environment. arXiv:1406.6222 (2014)
25. Zeitouni, O.: Random walks in random environment. In: Picard, J. (ed.) Lectures on Probability Theory and Statistics (Ecole D'ete de Probabilites de Saint-Flour XXXI-2001), Lecture Notes in Mathematics, vol. 1837, pp. 189–312, Springer, Berlin/Heidelberg (2004)

Part III
Population Growth Models in Varying and Random Environments

Chapter 5
Some Asymptotic Results for Strongly Critical Branching Processes with Immigration in Varying Environment

Márton Ispány

Mathematics Subject Classification (2000): 60J80, 60F17

5.1 Introduction

In the paper, $\mathbb{Z}_+$, $\mathbb{N}$, $\mathbb{R}$, and $\mathbb{R}_+$ denote the set of non-negative integers, positive integers, real numbers, and non-negative real numbers, respectively. Every random variable will be defined on a fixed probability space $(\Omega, \mathcal{A}, \mathsf{P})$.

Let $\{\xi_{k,j},\, \varepsilon_k : k, j \in \mathbb{N}\}$ be independent, non-negative, integer-valued random variables such that $\{\xi_{k,j} : j \in \mathbb{N}\}$ are identically distributed for each $k \in \mathbb{N}$. Define recursively

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{k,j} + \varepsilon_k \quad \text{for } k \in \mathbb{N}\,, \qquad X_0 = 0\,. \tag{5.1}$$

We can interpret X_k as the size of the kth generation of a population, where $\xi_{k,j}$ is the number of offsprings of the jth individual in the $(k-1)$st generation and ε_k is the number of immigrants contributing to the kth generation. We suppose that either the reproduction, or the immigration mechanism, or both need not stay the same from generation to generation. Assume that, for all $k \in \mathbb{N}$, the moments $m_k := \mathrm{E}(\xi_{k,1})$, $\lambda_k := \mathrm{E}(\varepsilon_k)$, $\sigma_k^2 := \mathsf{Var}(\xi_{k,1})$, and $b_k^2 := \mathsf{Var}(\varepsilon_k)$ are finite. The sequence $(X_k)_{k\in\mathbb{Z}_+}$ is called a branching process with immigration in varying environment (BPIVE) or a time-varying branching process with immigration.

M. Ispány (✉)
Faculty of Informatics, University of Debrecen, Pf. 12, H-4010 Debrecen, Hungary
e-mail: ispany.marton@inf.unideb.hu

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_5

The theory of branching stochastic models is widely applied in different fields of sciences, see, e.g., the survey Haccou et al. [5]. Most of the classical models have been studied in varying environment as well. See, e.g., the papers by Fearn [2], Jagers [8], Keiding and Nielsen [10] for the case of branching processes without immigration. The case of branching processes with immigration was considered, e.g., by Rahimov [12, Chap. III] and [13], Mitov and Omey [11].

For critical BPI in non-varying environment, i.e., when $m_k = 1$, $\lambda_k = \lambda$, $\sigma_k^2 = \sigma^2 > 0$, and $b_k^2 = b^2$ for all $k \in \mathbb{N}$, Wei and Winnicki [15] proved a functional limit theorem $\mathcal{X}^n \xrightarrow{\mathcal{L}} \mathcal{X}$ as $n \to \infty$, where $\mathcal{X}_t^n := n^{-1} X_{\lfloor nt \rfloor}$ for $t \in \mathbb{R}_+$, $n \in \mathbb{N}$, where $\lfloor x \rfloor$ denotes the lower integer-part of $x \in \mathbb{R}$, and $(\mathcal{X}_t)_{t \in \mathbb{R}_+}$ is a (nonnegative) diffusion process with initial value $\mathcal{X}_0 = 0$ defined as the unique strong solution to the stochastic differential equation (SDE)

$$\mathrm{d}\mathcal{X}_t = \lambda \, \mathrm{d}t + \sqrt{\sigma^2 \mathcal{X}_t^+} \, \mathrm{d}\mathcal{W}_t \,, \qquad t \in \mathbb{R}_+ \,, \tag{5.2}$$

where $(\mathcal{W}_t)_{t \in \mathbb{R}_+}$ is a standard Wiener process, and x^+ denotes the positive part of $x \in \mathbb{R}$. If the offspring variance σ^2 equals to 0 then a fluctuation limit theorem $\widetilde{\mathcal{X}}^n \xrightarrow{\mathcal{L}} b\mathcal{W}$ as $n \to \infty$ was proved, where $\widetilde{\mathcal{X}}_t^n := n^{-1/2}(X_{\lfloor nt \rfloor} - \mathrm{E}(X_{\lfloor nt \rfloor}))$ for $t \in \mathbb{R}_+$, $n \in \mathbb{N}$, and $(\mathcal{W}_t)_{t \in \mathbb{R}_+}$ is a standard Wiener process, see Ispány et al. [7, Theorem 2.2] in a more general setup where sequences of nearly critical BPI's were considered.

This paper presents an attempt to generalize the above mentioned limit theorems to varying environment when the parameters of the environment converge in some sense. To cover the most important cases, e.g., the periodically varying environment, we suppose that the parameters of BPIVE converge in the Cesaro sense. In Theorem 5.1 it is proved that $(\mathcal{X}^n)_{n \in \mathbb{N}}$ converges weakly to a diffusion process (5.3), which is similar to (5.2), where the diffusion coefficient is a linear combination of state and time values. For the fluctuation process $(\widetilde{\mathcal{X}}^n)_{n \in \mathbb{N}}$ the invariance principle remains hold, see Theorem 5.2. Note that Rahimov [13] derived various time changed Brownian motions as limit distributions in the case of fixed critical offspring distribution and regularly varying immigration processes. Finally, we recall that a branching process with immigration can be considered as a particular case of controlled branching processes (CBP), see, e.g., González and del Puerto [3, Remark 3(2)]. In [3] diffusion approximation was proved for an array of controlled branching processes under the conditions that the offspring and control means tend to be critical. Controlled branching processes in varying environment (CBPVE) were also investigated, see [3, Theorem 2], in the critical case, and see González et al. [4], in the uniformly sub- and supercritical cases, respectively. However, in these papers, only the offspring distribution may vary from one generation to another while the control distribution is supposed to be fixed in time. In present paper, time-varying immigration and thus time-varying control distribution, see Remark 5.3, is also allowed.

The paper is organized as follows. The main results together with the basic assumptions are stated in Sect. 5.2. Asymptotic formulas are derived for the first two moments in Sect. 5.3. The two limit theorems are proved in Sect. 5.4.

5.2 Main Results

Throughout the paper, we consider environments which converge in the Cesaro sense. Recall that a real sequence $(\lambda_k)_{k\in\mathbb{N}}$ converges in the Cesaro sense if there exists $\lambda \in \mathbb{R}$ such that, for the averages $\overline{\lambda}_n := n^{-1}\sum_{k=1}^{n}\lambda_k$, $n \in \mathbb{N}$, we have $\overline{\lambda}_n \to \lambda$ as $n \to \infty$. We denote the Cesaro convergence by $\lambda_k \rightsquigarrow \lambda$. Clearly, the convergence $\lambda_k \to \lambda$ implies the Cesaro convergence $\lambda_k \rightsquigarrow \lambda$. A BPIVE is called asymptotically critical if $m_n \to 1$ as $n \to \infty$. We suppose that the following stronger assumption A holds:

$$\sum_{k=1}^{\infty}|m_k - 1| < +\infty \qquad \text{and} \qquad \lambda_n \rightsquigarrow \lambda \quad \text{as} \quad n \to \infty\ .$$

If a BPIVE satisfies the first part of this assumption, i.e., the convergence of the time-varying offspring mean to 1 is fast enough, then it is called strongly critical. Moreover, we introduce the following assumption B for the variances of the offspring and immigration distributions:

$$\sigma_n^2 \rightsquigarrow \sigma^2 \geq 0 \qquad \text{and} \qquad n^{-1}b_n^2 \rightsquigarrow b^2 \geq 0 \quad \text{as} \quad n \to \infty\ .$$

Plainly, by the Toeplitz theorem, if $b_n^2 \rightsquigarrow b^2$ then $n^{-1}b_n^2 \rightsquigarrow 0$ as $n \to \infty$.

Example 5.1 A BPIVE possesses periodically varying immigration if there exists a period $s \in \mathbb{N}$ such that $\lambda_{k+s} = \lambda_k$ and $b_{k+s}^2 = b_k^2$ for all $k \in \mathbb{N}$. This clearly holds, e.g., if $\mathcal{L}(\varepsilon_{k+s}) = \mathcal{L}(\varepsilon_k)$ for all $k \in \mathbb{N}$, where $\mathcal{L}(\varepsilon)$ denotes the distribution of a random variable ε. Then, the second parts of assumptions A and B are fulfilled.

Example 5.2 Suppose that the offspring distributions follow a zero-modified geometric law such that $\xi = \eta\zeta$ where $\mathcal{L}(\eta)$ is a Bernoulli distribution with mean p and $\mathcal{L}(\zeta)$ is a geometric distribution with parameter p, where $p \in (0, 1)$. Then, $\mathrm{E}(\xi) = 1$ and $\mathsf{Var}(\xi) = 2(1-p)/p$, i.e., the process is critical and if p varies from generation to generation then the environment is varying. Such offspring distributions are investigated in [10, 11].

Let $\{U_k^n : k \in \mathcal{I}_n, n \in \mathbb{N}\}$ be a triangular system of random variables, where $\mathcal{I}_n$ is an index set for each $n \in \mathbb{N}$. Recall that $\{U_k^n : k \in \mathcal{I}_n, n \in \mathbb{N}\}$ satisfies the Lindeberg condition if

$$\sum_{k\in\mathcal{I}_n} \mathrm{E}\big(\big(U_k^n - \mathrm{E}(U_k^n)\big)^2 \mathbb{1}_{\{|U_k^n-\mathrm{E}(U_k^n)|>\theta\}}\big) \to 0$$

as $n \to \infty$ for all $\theta > 0$.

A function $f : \mathbb{R}_+ \to \mathbb{R}$ is called càdlàg if it is right continuous with left limits. Let $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ denote the space of all real-valued càdlàg functions on $\mathbb{R}_+$. Let $\mathcal{D}_\infty$ denote the Borel σ-field in $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ for the metric defined in (16.4) in Billingsley [1] (with this metric $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$ is a complete and separable metric space). For stochastic processes $(\mathcal{U}_t)_{t\in\mathbb{R}_+}$ and $(\mathcal{U}_t^n)_{t\in\mathbb{R}_+}$, $n \in \mathbb{N}$, with càdlàg paths we write $\mathcal{U}^n \xrightarrow{\mathcal{L}} \mathcal{U}$ if the distribution of $\mathcal{U}^n$ on the space $(\mathbb{D}(\mathbb{R}_+, \mathbb{R}), \mathcal{D}_\infty)$ converges weakly to the distribution of $\mathcal{U}$ on the space $(\mathbb{D}(\mathbb{R}_+, \mathbb{R}), \mathcal{D}_\infty)$ as $n \to \infty$.

The first main theorem of the paper is the following analogue of the result of Wei and Winnicki.

Theorem 5.1 *Let $(X_k)_{k\in\mathbb{Z}_+}$ be a BPIVE. Suppose that assumptions A and B hold, and the triangular systems $\{n^{-1}\xi_{k,j} : k, j = 1, \ldots, n, n \in \mathbb{N}\}$ and $\{n^{-1}\varepsilon_k : k = 1, \ldots, n, n \in \mathbb{N}\}$ satisfy the Lindeberg condition. For each $n \in \mathbb{N}$, introduce the random step function $\mathcal{X}_t^n := n^{-1}X_{\lfloor nt\rfloor}$, $t \in \mathbb{R}_+$. Then, weakly in the Skorokhod space $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$,*

$$\mathcal{X}^n \xrightarrow{\mathcal{L}} \mathcal{X} \qquad \text{as} \qquad n \to \infty\,,$$

where $(\mathcal{X}_t)_{t\in\mathbb{R}_+}$ is the unique strong solution to the SDE

$$\mathrm{d}\mathcal{X}_t = \lambda\,\mathrm{d}t + \sqrt{\sigma^2\mathcal{X}_t^+ + b^2 t}\,\mathrm{d}\mathcal{W}_t\,, \qquad t \in \mathbb{R}_+\,, \tag{5.3}$$

with initial value $\mathcal{X}_0 = 0$, where $(\mathcal{W}_t)_{t\in\mathbb{R}_+}$ is a standard Wiener process.

Remark 5.1 The SDE (5.3) has a unique strong solution $(\mathcal{X}_t^x)_{t\in\mathbb{R}_+}$ for all initial values $\mathcal{X}_0^x = x \in \mathbb{R}$. Indeed, since $|\sqrt{\sigma^2 x + b^2 t} - \sqrt{\sigma^2 y + b^2 t}| \le \sigma\sqrt{|x-y|}$ for all $x, y, t \in \mathbb{R}_+$, the coefficient functions $\mathbb{R}_+ \times \mathbb{R} \ni (t, x) \mapsto \lambda$ and $\mathbb{R}_+ \times \mathbb{R} \ni (t, x) \mapsto \sqrt{\sigma^2 x^+ + b^2 t}$ satisfy conditions of part (ii) of Theorem 3.5 in Chap. IX in Revuz and Yor [14] or the conditions of Proposition 5.2.13 in Karatzas and Shreve [9]. Further, by the comparison theorem (see, e.g., Revuz and Yor [14, Theorem 3.7, Chap. IX]), if the initial value $\mathcal{X}_0^x = x$ is non-negative, then $\mathcal{X}_t^x$ is nonnegative for all $t \in \mathbb{R}_+$ with probability one. Hence, $\mathcal{X}_t^+$ may be replaced by $\mathcal{X}_t$ under the square root in (5.3).

Remark 5.2 In the case of $\sigma^2 > 0$ and $b^2 = 0$ the SDE (5.3) is simplified to the SDE (5.2) and the unique strong solution of the SDE (5.2) is known as a squared Bessel process, a squared-root process, or a Cox-Ingersoll-Ross (CIR) process. In the case of $\sigma^2 = 0$ and $b^2 > 0$ the unique strong solution of the SDE (5.3) is given by $\mathcal{X}_t = \lambda t + \int_0^t \sqrt{b^2 s}\,\mathrm{d}\mathcal{W}_s$, $t \in \mathbb{R}_+$. We remark that $(\mathcal{X}_t)_{t\in\mathbb{R}_+}$ is a continuous

Gaussian process with independent (but not stationary) increments. Its mean and covariance functions have the forms $\mathrm{E}(\mathcal{X}_t) = \lambda t$, $t \in \mathbb{R}_+$, and $\mathsf{Cov}(\mathcal{X}_t, \mathcal{X}_s) = T(t \wedge s)$, $t, s \in \mathbb{R}_+$, where $T(t) := 2^{-1}(bt)^2$, $t \in \mathbb{R}_+$.

Remark 5.3 Define the random variables $Z_k := X_k - \varepsilon_k$, $k \in \mathbb{N}$, and $\phi_k(\ell) := \ell + \varepsilon_k$, $\ell \in \mathbb{Z}_+$, $k \in \mathbb{N}$. Then, the sequence $(Z_k)_{k\in\mathbb{N}}$ satisfies the recursion

$$Z_{k+1} = \sum_{j=1}^{\phi_k(Z_k)} \xi_{k+1,j} \quad \text{for } k \in \mathbb{N}\,, \qquad Z_1 = 0\,,$$

and thus the stochastic process $(Z_k)_{k\in\mathbb{N}}$ is a CBP with random control functions $(\phi_k)_{k\in\mathbb{N}}$, see, e.g., [3, 4]. If the first two moments of the immigration distributions are time-homogeneous, i.e., $\lambda_k = \lambda$ and $b_k^2 = b^2$ for all $k \in \mathbb{N}$, then the process $(Z_k)_{k\in\mathbb{N}}$ satisfies the assumptions of [3, Theorem 2] with the choice $m = 1$, $\alpha = 0$ and $\gamma = \lambda$. It is easy to see that the asymptotic behaviors of $(X_k)_{k\in\mathbb{N}}$ and $(Z_k)_{k\in\mathbb{N}}$ coincide, they converge (under the same normalization) to the same limit process defined by (5.2). Hence, both Theorem 5.1 and [3, Theorem 2] hold equally for a same wide class of branching processes, although the methods of their proofs are different: a martingale limit theorem is applied in the proof of Theorem 5.1, while an operator semigroup convergence theorem is used in [3]. However, [3, Theorem 2] does not cover the case of time-varying control which occurs when a time-varying immigration is involved to the model, while Theorem 5.1 of this paper is capable to handle this case as well. The extension of the results of this paper to the case of CBP with time-varying random control remains a task for the future.

If $\sigma^2 = b^2 = 0$ in assumption B then the limit in Theorem 5.1 will be the deterministic process $\mathcal{X}_t = \lambda t$, $t \in \mathbb{R}_+$. In this case, let us introduce the following assumption C for the variances of the offspring and immigration distributions:

$$n\sigma_n^2 \rightsquigarrow \sigma^2 \geq 0 \qquad \text{and} \qquad b_n^2 \rightsquigarrow b^2 \geq 0 \quad \text{as} \quad n \to \infty\,.$$

Clearly, this assumption implies assumption B with $\sigma^2 = b^2 = 0$. The second main result of the paper, which is an extension of [7, Theorem 2.2] for varying environment, is the following fluctuation limit theorem.

Theorem 5.2 *Let $(X_k)_{k\in\mathbb{Z}_+}$ be a BPIVE. Suppose that assumptions A and C hold, and the triangular systems $\{n^{-1/2}\xi_{k,j} : k, j = 1, \ldots, n, n \in \mathbb{N}\}$ and $\{n^{-1/2}\varepsilon_k : k = 1, \ldots, n, n \in \mathbb{N}\}$ satisfy the Lindeberg condition. For each $n \in \mathbb{N}$, introduce the random step function $\widetilde{\mathcal{X}}_t^n := n^{-1/2}\widetilde{X}_{\lfloor nt\rfloor}$, $t \in \mathbb{R}_+$, where $\widetilde{X}_k := X_k - \mathrm{E}(X_k)$, $k \in \mathbb{Z}_+$. Then, weakly in the Skorokhod space $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$,*

$$\widetilde{\mathcal{X}}^n \xrightarrow{\mathcal{L}} \tilde{\sigma}\mathcal{W} \qquad as \qquad n \to \infty\,, \tag{5.4}$$

where $\tilde{\sigma}^2 := \sigma^2\lambda + b^2$ and $(\mathcal{W}_t)_{t\in\mathbb{R}_+}$ is a standard Wiener process.

In the proofs of the main theorems and propositions we often apply the associated martingale differences and some basic regression equations. For $k \in \mathbb{Z}_+$, let $\mathcal{F}_k$ denote the σ-algebra generated by $X_0, X_1, \dots, X_k$. Then, by (5.1), we have the conditional expectation

$$\mathrm{E}(X_k \,|\, \mathcal{F}_{k-1}) = m_k X_{k-1} + \lambda_k \,, \qquad k \in \mathbb{N} \,. \tag{5.5}$$

Clearly,

$$M_k := X_k - \mathrm{E}(X_k \,|\, \mathcal{F}_{k-1}) = X_k - m_k X_{k-1} - \lambda_k \,, \quad k \in \mathbb{N} \,, \tag{5.6}$$

defines a martingale difference sequence $(M_k)_{k\in\mathbb{N}}$ with respect to the filtration $(\mathcal{F}_k)_{k\in\mathbb{Z}_+}$. Moreover, we have the recursion (regression equation)

$$X_k = m_k X_{k-1} + M_k + \lambda_k \quad \text{for } k \in \mathbb{N} \,, \qquad X_0 = 0 \,. \tag{5.7}$$

For the centered process $(\widetilde{X}_k)_{k\in\mathbb{N}}$, by (5.5) and (5.7), we have the recursion

$$\widetilde{X}_k = m_k \widetilde{X}_{k-1} + M_k \quad \text{for } k \in \mathbb{N} \,, \qquad \widetilde{X}_0 = 0 \,. \tag{5.8}$$

5.3 Estimations of Moments

First, we need formulas for the martingale difference sequence $(M_k)_{k\in\mathbb{N}}$.

Lemma 5.1 *For all $k \in \mathbb{N}$, the decomposition*

$$M_k = \sum_{j=1}^{X_{k-1}} (\xi_{k,j} - m_k) + (\varepsilon_k - \lambda_k) =: \zeta_k^1 + \zeta_k^2 \tag{5.9}$$

holds which implies $\mathrm{E}(M_k \,|\, \mathcal{F}_{k-1}) = 0$ *and* $\mathrm{E}(M_k^2 \,|\, \mathcal{F}_{k-1}) = \sigma_k^2 X_{k-1} + b_k^2$.

Proof By (5.1) and (5.6) we have (5.9). The second part of the lemma follows from the fact that the random variables $\{\xi_{k,j} - m_k, \varepsilon_k - \lambda_k : j \in \mathbb{N}\}$ are independent of each others, independent of $\mathcal{F}_{k-1}$, and have zero mean. □

In the next proposition we describe the asymptotic behavior of the first two moments of the process $(X_k)_{k\in\mathbb{Z}_+}$. In the proof of the proposition we apply the following simple lemma on Toeplitz summation. In the sequel, C denotes various constants that depend on the context.

Lemma 5.2 *Let $\{a_{n,k} : k = 1, \dots, n, n \in \mathbb{N}\}$ be a real kernel such that $a_{n,k} \to 0$ as $n \to \infty$ for each $k \in \mathbb{N}$, $\sum_{k=1}^n a_{n,k} \to a \in \mathbb{R}$ as $n \to \infty$, and* $\sup_{n\in\mathbb{N}} \sum_{k=1}^n k|a_{n,k+1} - a_{n,k}| < +\infty$, *where $a_{n,n+1} := 0$. Moreover, let $(\lambda_k)_{k\in\mathbb{N}}$ be a real sequence such that $\lambda_k \rightsquigarrow \lambda$ as $k \to \infty$. Then $\sum_{k=1}^n a_{n,k}\lambda_k \to a\lambda$ as $n \to \infty$.*

Proof Let $\Lambda_n := \sum_{k=1}^{n} \lambda_k = n\overline{\lambda}_n$, $n \in \mathbb{N}$. By summation by parts we have

$$\sum_{k=1}^{n} a_{n,k}\lambda_k = a_{n,n}\Lambda_n - \sum_{k=1}^{n-1}(a_{n,k+1} - a_{n,k})\Lambda_k = \sum_{k=1}^{n} b_{n,k}\overline{\lambda}_k \ ,$$

where $b_{n,k} := k(a_{n,k} - a_{n,k+1})$, $k = 1, \ldots, n-1$, $b_{n,n} := na_{n,n}$, $n \in \mathbb{N}$. Since $b_{n,k} \to 0$ as $n \to \infty$ for each $k \in \mathbb{N}$ and $\sum_{k=1}^{n} b_{n,k} = \sum_{k=1}^{n} a_{n,k} \to a$ as $n \to \infty$, the Toeplitz theorem implies the statement of the lemma. ☐

Proposition 5.1 *Let $(X_k)_{k\in\mathbb{Z}_+}$ be a BPIVE and suppose that assumption A holds. (i) Then $n^{-1}\mathrm{E}(X_n) \to \lambda$ as $n \to \infty$. (ii) Suppose, in addition, that assumption B holds. Then $n^{-2}\mathsf{Var}(X_n) \to (\sigma^2\lambda + b^2)/2$ as $n \to \infty$. (iii) Finally, if assumption C holds also then $n^{-1}\mathsf{Var}(X_n) \to \sigma^2\lambda + b^2$ as $n \to \infty$.*

Proof

(i) Introduce the non-negative sequence $x_k := \mathrm{E}(X_k)$, $k \in \mathbb{Z}_+$. Then, by (5.7), we have the recursion $x_k = m_k x_{k-1} + \lambda_k$, $k \in \mathbb{N}$, with initial value $x_0 = 0$. Clearly, for the recursion $y_k = y_{k-1} + \lambda_k$, $k \in \mathbb{N}$, with initial value $y_0 := 0$, we obtain that $n^{-1}y_n = n^{-1}\sum_{k=1}^{n}\lambda_k \to \lambda$ as $n \to \infty$. On the other hand, for the real sequence $z_k := x_k - y_k$, $k \in \mathbb{Z}_+$, the recursion $z_k = m_k z_{k-1} + (m_k - 1)y_{k-1}$, $k \in \mathbb{N}$, can be derived with initial value $z_0 = 0$. Hence, we have $z_k = \sum_{j=1}^{k}\prod_{i=j+1}^{k} m_i(m_j - 1)y_{j-1}$, $k \in \mathbb{N}$. Since $\prod_{i=k}^{\ell} m_i \le \exp\left\{\sum_{i=k}^{\ell}|m_i - 1|\right\} \le \exp\left\{\sum_{i=1}^{\infty}|m_i - 1|\right\} < +\infty$ for all $0 < k \le \ell$ by assumption A and $\{n^{-1}y_n\}_{n\in\mathbb{N}}$ is a bounded sequence, we have

$$n^{-1}|z_n| \le n^{-1}\sum_{j=1}^{n}\prod_{i=j+1}^{n} m_i|m_j - 1|y_{j-1} \le Cn^{-1}\sum_{j=1}^{n} j|m_j - 1| \to 0 \tag{5.10}$$

as $n \to \infty$ by Kronecker's lemma. Thus, $|n^{-1}x_n - \lambda| \le |n^{-1}y_n - \lambda| + n^{-1}|z_n| \to 0$ as $n \to \infty$ which proves (i).

(ii) Introduce the non-negative sequence $\tilde{x}_k := \mathrm{E}(\widetilde{X}_k^2) = \mathsf{Var}(X_k)$, $k \in \mathbb{Z}_+$. Then, by (5.8), we have the recursion $\tilde{x}_k = m_k^2\tilde{x}_{k-1} + \mathrm{E}(M_k^2)$, $k \in \mathbb{N}$, with initial value $\tilde{x}_0 = 0$. Define the recursion $\tilde{y}_k = \tilde{y}_{k-1} + \mathrm{E}(M_k^2)$, $k \in \mathbb{N}$, with $\tilde{y}_0 := 0$. Consider the sequence $\psi_k := k^{-1}\mathrm{E}(M_k^2)$, $k \in \mathbb{N}$. Then, by Lemma 5.1, $\psi_k = \sigma_k^2 k^{-1}\mathrm{E}(X_{k-1}) + k^{-1}b_k^2$. Assumption B and (i) of Proposition 5.1, by the Toeplitz theorem, imply that $\psi_n \rightsquigarrow \sigma^2\lambda + b^2$ as $n \to \infty$. The kernel $\{a_{n,k} : k = 1, \ldots, n, n \in \mathbb{N}\}$ defined by $a_{n,k} := n^{-2}k$ satisfies the assumptions of Lemma 5.2 with $\sum_{k=1}^{n} a_{n,k} \to 1/2$ as $n \to \infty$. Thus, by Lemma 5.2, we have

$$n^{-2}\tilde{y}_n = n^{-2}\sum_{j=1}^{n}\mathrm{E}(M_j^2) = \sum_{j=1}^{n} a_{n,j}\psi_j \to \frac{1}{2}(\sigma^2\lambda + b^2) \tag{5.11}$$

as $n \to \infty$. For the sequence $\tilde{z}_k := \tilde{x}_k - \tilde{y}_k$, $k \in \mathbb{Z}_+$, we obtain the recursion $\tilde{z}_k = m_k^2 \tilde{z}_{k-1} + (m_k^2 - 1)\tilde{y}_{k-1}$, $k \in \mathbb{N}$, with $\tilde{z}_0 = 0$. Thus, similarly to (5.10),

$$n^{-2}|\tilde{z}_n| \le n^{-2} \sum_{j=1}^{n} \prod_{i=j+1}^{n} m_i^2 |m_j^2 - 1| \tilde{y}_{j-1} \le Cn^{-2} \sum_{j=1}^{n} j^2 |m_j^2 - 1| \to 0$$

as $n \to \infty$ by Kronecker's lemma, since $\sum_{k=1}^{\infty} |m_k^2 - 1|$ is finite by assumption A. We have the assertion (ii) by $|n^{-2}\tilde{x}_n - (\sigma^2\lambda + b^2)/2| \le |n^{-2}\tilde{y}_n - (\sigma^2\lambda + b^2)/2| + n^{-2}|\tilde{z}_n| \to 0$ as $n \to \infty$.

(iii) Consider the sequences $(\tilde{x}_k)_{k\in\mathbb{Z}_+}$, $(\tilde{y}_k)_{k\in\mathbb{Z}_+}$ defined in the proof of (ii). By Lemma 5.1 and the Toeplitz theorem we have

$$n^{-1}\tilde{y}_n = n^{-1} \sum_{k=1}^{n} \mathrm{E}(M_k^2) = n^{-1} \sum_{k=1}^{n} \left(k\sigma_k^2 (k^{-1}\mathrm{E}(X_{k-1}) + b_k^2\right) \to \sigma^2\lambda + b^2 \tag{5.12}$$

as $n \to \infty$, since $\{n^{-1}k\sigma_k^2 : k = 1, \ldots, n, n \in \mathbb{N}\}$ is a Toeplitz kernel such that, as $n \to \infty$, $n^{-1}\sum_{k=1}^{n} k\sigma_k^2 \to \sigma^2$ by assumption C. Then, (iii) can be proved by similar arguments to the proofs of (i) and (ii).

□

Proposition 5.1 implies the following good bounds for the first two moments of the process $(X_k)_{k\in\mathbb{Z}_+}$. Under assumption A we have $\mathrm{E}(X_n) = O(n)$, under assumptions A and B we have $\mathsf{Var}(X_n) = O(n^2)$, finally, under assumptions A and C we have $\mathsf{Var}(X_n) = O(n)$. In the following proposition, we prove that these bounds hold uniformly as well.

Proposition 5.2 *Let* $(X_k)_{k\in\mathbb{Z}_+}$ *be a BPIVE. For all* $n \in \mathbb{N}$, *define* $V_n := \max_{1\le k\le n} X_k$ *and* $W_n := \max_{1\le k\le n}(X_k - \mathrm{E}(X_k))^2$. *(i) If assumptions A and B hold then* $\mathrm{E}(V_n) = O(n)$ *and* $\mathrm{E}(W_n) = O(n^2)$. *(ii) If assumptions A and C hold then* $\mathrm{E}(V_n) = O(n)$ *and* $\mathrm{E}(W_n) = O(n)$.

Proof

(i) Define the random variables $\widetilde{Y}_k := \sum_{j=1}^{k} M_j$, $k \in \mathbb{N}$. Then, $(\widetilde{Y}_k)_{k\in\mathbb{N}}$ is a martingale with respect to the filtration $(\mathcal{F}_k)_{k\in\mathbb{Z}_+}$. By Doob's inequality we have

$$\mathrm{E}\Big(\max_{1\le k\le n} \widetilde{Y}_k^2\Big) \le 4\mathrm{E}\Big(\sum_{j=1}^{n} M_j\Big)^2 = 4\sum_{j=1}^{n} \mathrm{E}\left(M_j^2\right) \tag{5.13}$$

and (5.11) implies that $\mathrm{E}(\max_{1\le k\le n} \widetilde{Y}_k^2) = O(n^2)$. Clearly, $\widetilde{Y}_k = \widetilde{Y}_{k-1} + M_k$, $k \in \mathbb{N}$, with initial value $\widetilde{Y}_0 = 0$. Thus, for the random sequence $\widetilde{Z}_k := \widetilde{X}_k - \widetilde{Y}_k$, $k \in \mathbb{N}$, by (5.8) we obtain the recursion $\widetilde{Z}_k = m_k \widetilde{Z}_{k-1} + (m_k - 1)\widetilde{Y}_{k-1}$, $k \in \mathbb{N}$, with $\widetilde{Z}_0 = 0$. Hence, similarly to (5.10), assumption A implies that, for all

$k = 1, \ldots, n$,

$$|\widetilde{Z}_k| = \left| \sum_{j=1}^{k} \prod_{i=j+1}^{k} m_i(m_j - 1)\widetilde{Y}_{j-1} \right| \le C \max_{1 \le k \le n} |\widetilde{Y}_k| \,.$$

By inequality $\widetilde{X}_k^2 \le 2(\widetilde{Y}_k^2 + \widetilde{Z}_k^2)$, we have $\mathrm{E}(\max_{1\le k\le n} \widetilde{X}_k^2) = O(n^2)$ which is the second statement of (i). Moreover, this implies $\mathrm{E}(\max_{1\le k\le n} |\widetilde{X}_k|) = O(n)$ by Lyapunov's inequality. Since $\mathrm{E}(V_n) \le \max_{1\le k\le n} \mathrm{E}(X_k) + \mathrm{E}(\max_{1\le k\le n} |\widetilde{X}_k|)$, by (i) of Proposition 5.1, we have the first statement of (i).

(ii) The proof is similar to part (i) bearing in mind that in inequality (5.13) we have $\mathrm{E}(\max_{1\le k\le n} \widetilde{Y}_k^2) = O(n)$ by (5.12).

□

5.4 Proofs of the Main Theorems

In the proofs of the main theorems we apply the following two lemmas of calculus several times.

Lemma 5.3 *Let $f_n : \mathbb{R}_+ \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of non-decreasing functions converging pointwise to the continuous function $f : \mathbb{R}_+ \to \mathbb{R}$. Then, for all $T > 0$, $\sup_{t\in[0,T]} |f_n(t) - f(t)| \to 0$ as $n \to \infty$.*

Lemma 5.4 *Let $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ be two Cesaro convergent sequences of non-negative real numbers. Then $n^{-2} \sum_{k=1}^{n} a_k b_k \to 0$ as $n \to \infty$.*

Proof of Theorem 5.1 We apply Theorem 5.3 (see Appendix) with the choice $U_k^n := n^{-1}(X_k - X_{k-1})$, $k, n \in \mathbb{N}$, $\mathcal{F}_k^n := \mathcal{F}_k$, $k \in \mathbb{Z}_+, n \in \mathbb{N}$, and $\beta(t,x) := \lambda$, $\gamma(t,x) := (\sigma^2 x + b^2 t)^{1/2}$, $t, x \in \mathbb{R}_+$. Then $\mathcal{U}^n = \mathcal{X}^n$, $n \in \mathbb{N}$, and $\mathcal{U} = \mathcal{X}$ with $\mathcal{U}_0 = 0$, where $(\mathcal{X}_t)_{t\in\mathbb{R}_+}$ is defined by (5.3). We have to prove for all $T > 0$ that

$$\sup_{t\in[0,T]} \left| \sum_{k=1}^{\lfloor nt \rfloor} \mathrm{E}\big(U_k^n \,|\, \mathcal{F}_{k-1}\big) - \lambda t \right| \xrightarrow{\mathrm{P}} 0 \,, \tag{5.14}$$

$$\sup_{t\in[0,T]} \left| \sum_{k=1}^{\lfloor nt \rfloor} \mathrm{Var}\big(U_k^n \,|\, \mathcal{F}_{k-1}\big) - \int_0^t \big(\sigma^2 \mathcal{X}_s^n + b^2 s\big)\, \mathrm{d}s \right| \xrightarrow{\mathrm{P}} 0 \tag{5.15}$$

as $n \to \infty$, where $\xrightarrow{\mathrm{P}}$ denotes convergence in probability, and (5.40) holds for the pair $(U_k^n)_{k\in\mathbb{N}}$, $(\mathcal{F}_k)_{k\in\mathbb{Z}_+}$, $n \in \mathbb{N}$.

Since $\mathrm{E}(U_k^n \,|\, \mathcal{F}_{k-1}) = n^{-1}((m_k - 1)X_{k-1} + \lambda_k)$, $k, n \in \mathbb{N}$, in order to prove (5.14) we have to show that

$$\sup_{t\in[0,T]} \left| n^{-1} \sum_{k=1}^{\lfloor nt \rfloor} \lambda_k - \lambda t \right| \to 0 \,, \qquad n^{-1} \sum_{k=1}^{\lfloor nT \rfloor} |m_k - 1| X_{k-1} \xrightarrow{\mathrm{P}} 0 \tag{5.16}$$

as $n \to \infty$. Clearly, the first statement of (5.16) follows from assumption A by Lemma 5.3. One can see that it is enough to prove the second statement of (5.16) for $T = 1$. By Markov's inequality and (i) of Proposition 5.1, it is enough to see that

$$n^{-1} \sum_{k=1}^{n} |m_k - 1| \mathrm{E}(X_{k-1}) \le C n^{-1} \sum_{k=1}^{n} k |m_k - 1| \to 0$$

as $n \to \infty$ which follows by Kronecker's lemma.

By Lemma 5.1 we obtain $\mathsf{Var}(U_k^n \,|\, \mathcal{F}_{k-1}) = n^{-2}\mathrm{E}(M_k^2 \,|\, \mathcal{F}_{k-1}) = n^{-2}(\sigma_k^2 X_{k-1} + b_k^2)$, $k, n \in \mathbb{N}$. Moreover, for $t \in \mathbb{R}_+$, $n \in \mathbb{N}$,

$$\int_0^t \mathcal{X}_s^n \,\mathrm{d}s = n^{-2} \sum_{k=1}^{\lfloor nt \rfloor} X_{k-1} + n^{-2}(nt - \lfloor nt \rfloor) X_{\lfloor nt \rfloor} \,.$$

Thus, in order to prove (5.15) it is enough to show that

$$\sup_{t\in[0,T]} n^{-2} \left| \sum_{k=1}^{\lfloor nt \rfloor} \sigma_k^2 X_{k-1} - \sigma^2 \sum_{k=1}^{\lfloor nt \rfloor} X_{k-1} \right| \xrightarrow{\mathrm{P}} 0 \,, \tag{5.17}$$

$$n^{-2} V_{\lfloor nT \rfloor} \xrightarrow{\mathrm{P}} 0 \,, \qquad \sup_{t\in[0,T]} \left| n^{-2} \sum_{k=1}^{\lfloor nt \rfloor} b_k^2 - 2^{-1} b^2 t^2 \right| \to 0 \tag{5.18}$$

as $n \to \infty$. Recall that for any real sequence $(a_k)_{k\in\mathbb{N}}$ with partial sum $A_n := \sum_{k=1}^{n} a_k$, $n \in \mathbb{N}$, we have, by summation by parts,

$$\sum_{k=1}^{n} a_k X_{k-1} = A_n X_n - \sum_{k=1}^{n} A_k (X_k - X_{k-1}) \,,$$

since $X_0 = 0$. Using this formula for the sequences $(\sigma_k^2)_{k\in\mathbb{N}}$ and $(1)_{k\in\mathbb{N}}$, since by (5.7)

$$X_k - X_{k-1} = (m_k - 1) X_{k-1} + \lambda_k + M_k \,, \qquad k \in \mathbb{N} \,, \tag{5.19}$$

we have

$$\left|\sum_{k=1}^{n}(\sigma_k^2-\sigma^2)X_{k-1}\right| \leq |d_n|X_n + |S_n| + \sum_{k=1}^{n}|m_k-1||d_k|X_{k-1} + \sum_{k=1}^{n}|d_k|\lambda_k \ ,$$

where $d_n := \sum_{k=1}^{n}\sigma_k^2 - n\sigma^2$ and $S_n := \sum_{k=1}^{n} d_k M_k$, $n \in \mathbb{N}$. Hence, since it is enough to prove (5.17) for $T = 1$, we have to show that, as $n \to \infty$,

$$n^{-2}\max_{1\leq k\leq n}|d_k|X_k \xrightarrow{\mathrm{P}} 0 \ , \qquad n^{-2}\max_{1\leq k\leq n}|S_k| \xrightarrow{\mathrm{P}} 0 \ , \tag{5.20}$$

$$n^{-2}\sum_{k=1}^{n}|m_k-1||d_k|X_{k-1} \xrightarrow{\mathrm{P}} 0 \ , \qquad n^{-2}\sum_{k=1}^{n}|d_k|\lambda_k \to 0 \ . \tag{5.21}$$

By assumption B there exists $D \in \mathbb{R}_+$ such that $|d_n| < Dn$ for all $n \in \mathbb{N}$, and for all $\epsilon > 0$ there exists $N_\epsilon \in \mathbb{N}$ such that $|d_n| < \epsilon n$ if $n > N_\epsilon$. Thus, by (i) of Proposition 5.2, if $n > N_\epsilon$ then we have

$$\overline{\lim_{n\to\infty}}\ \mathrm{E}\left(n^{-2}\max_{1\leq k\leq n}|d_k|X_k\right) \leq \overline{\lim_{n\to\infty}}\left(n^{-1}D\mathrm{E}\left(V_{N_\epsilon}\right) + \epsilon n^{-1}\mathrm{E}\left(V_n\right)\right) \leq C\epsilon \ ,$$

where C is a constant which does not depend on ϵ. Then, letting $\epsilon \to 0$, we obtain the first assertion of (5.20) by Markov's inequality. In order to prove the second assertion of (5.20) we apply the Doob inequality for the martingale $(S_n)_{n\in\mathbb{N}}$ (with respect to the filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$) and we have

$$\mathrm{E}\Big(\max_{1\leq k\leq n} S_k^2\Big) \leq 4\mathrm{E}\Big(\sum_{k=1}^{n} d_k M_k\Big)^2 = 4\sum_{k=1}^{n} d_k^2\mathrm{E}\left(M_k^2\right) \ .$$

Thus, by (5.11), if $n > N_\epsilon$ then we obtain

$$\overline{\lim_{n\to\infty}}\ \mathrm{E}\Big(n^{-4}\max_{1\leq k\leq n} S_k^2\Big) \leq \overline{\lim_{n\to\infty}}\Big(\frac{4D^2}{n^2}\sum_{k=1}^{N_\epsilon}\mathrm{E}\left(M_k^2\right) + \frac{4\varepsilon^2}{n^2}\sum_{k=1}^{n}\mathrm{E}\left(M_k^2\right)\Big) = 2\varepsilon^2(\sigma^2\lambda + b^2)$$

and letting $\epsilon \to 0$ we obtain the second assertion of (5.20) by Markov's and Lyapunov's inequalities. The first assertion of (5.21) follows by Markov's inequality and (i) of Proposition 5.1 since

$$n^{-2}\sum_{k=1}^{n}|m_k-1||d_k|\mathrm{E}(X_{k-1}) \leq Cn^{-2}\sum_{k=1}^{n}k^2|m_k-1| \to 0$$

as $n \to \infty$ by the Kronecker lemma. Finally, the second assertion of (5.21) follows by letting $\epsilon \to 0$ in the inequality

$$\overline{\lim_{n\to\infty}}\left(n^{-2}\sum_{k=1}^{n}|d_k|\lambda_k\right) \le \overline{\lim_{n\to\infty}}\left(\frac{D}{n}\sum_{k=1}^{N_\epsilon}\lambda_k + \frac{\epsilon}{n}\sum_{k=1}^{n}\lambda_k\right) = \epsilon\lambda \ ,$$

where we used assumption A. Clearly, the first assertion of (5.18) follows from (i) of Proposition 5.2 by Markov's inequality. Since, for all $t \in \mathbb{R}_+$,

$$n^{-2}\sum_{k=1}^{\lfloor nt\rfloor} b_k^2 = \frac{\lfloor nt\rfloor^2}{n^2}\sum_{k=1}^{\lfloor nt\rfloor} a_{\lfloor nt\rfloor,k}k^{-1}b_k^2 \to 2^{-1}b^2t^2$$

as $n \to \infty$, where the kernel $\{a_{n,k} : k = 1,\dots,n, n \in \mathbb{N}\}$ is defined in the proof of Proposition 5.1 (ii), by assumption B and the Toeplitz theorem, Lemma 5.3 implies the second assertion of (5.18).

In order to prove (5.40) for the pair $(U_k^n)_{k\in\mathbb{N}}, (\mathcal{F}_k)_{k\in\mathbb{Z}_+}$, $n \in \mathbb{N}$, we note that $(U_k^n)^2 \le 3n^{-2}((m_k-1)^2X_{k-1}^2 + \lambda_k^2 + M_k^2)$ by (5.19). Thus, it is enough to show that

$$n^{-2}\sum_{k=1}^{\lfloor nT\rfloor}(m_k-1)^2X_{k-1}^2 \xrightarrow{\mathrm{P}} 0 \ , \qquad n^{-2}\sum_{k=1}^{\lfloor nT\rfloor}\lambda_k^2 \to 0 \ , \tag{5.22}$$

$$n^{-2}\sum_{k=1}^{\lfloor nT\rfloor}\mathrm{E}\big(M_k^2\mathbb{1}_{\{|U_k^n|>\theta\}}\,\big|\,\mathcal{F}_{k-1}\big) \xrightarrow{\mathrm{P}} 0 \quad \text{for all } \theta > 0 \tag{5.23}$$

as $n \to \infty$. Clearly, for (5.22), one can suppose that $T = 1$. By Markov's inequality we have the first statement of (5.22) since the Kronecker lemma implies

$$n^{-2}\sum_{k=1}^{n}(m_k-1)^2\mathrm{E}\left(X_{k-1}^2\right) \le Cn^{-2}\sum_{k=1}^{n}k^2(m_k-1)^2 \to 0$$

as $n \to \infty$, where we applied Proposition 5.1 and $\sum_{k=1}^{\infty}(m_k-1)^2 < +\infty$. Lemma 5.4 implies the second statement of (5.22) from assumption A. In order to prove (5.23) we note that

$$\mathbb{1}_{\{|Y+Z|>\theta\}} \le \mathbb{1}_{\{|Y|>\theta/2\}} + \mathbb{1}_{\{|Z|>\theta/2\}} \tag{5.24}$$

for any pair of random variables Y and Z. Hence, it is enough to show that

$$n^{-2}\sum_{k=1}^{\lfloor nT\rfloor}\mathrm{E}\big(M_k^2\mathbb{1}_{\{|(m_k-1)X_{k-1}|>\theta n\}}\,\big|\,\mathcal{F}_{k-1}\big) \xrightarrow{\mathrm{P}} 0 \ , \tag{5.25}$$

$$n^{-2}\sum_{k=1}^{\lfloor nT\rfloor}\mathrm{E}\big(M_k^2\mathbb{1}_{\{|\lambda_k|>\theta n\}}\,\big|\,\mathcal{F}_{k-1}\big)\xrightarrow{\mathrm{P}}0\,, \tag{5.26}$$

$$n^{-2}\sum_{k=1}^{\lfloor nT\rfloor}\mathrm{E}\big(M_k^2\mathbb{1}_{\{|M_k|>\theta n\}}\,\big|\,\mathcal{F}_{k-1}\big)\xrightarrow{\mathrm{P}}0 \tag{5.27}$$

as $n\to\infty$ for all $\theta>0$. Clearly, one can also suppose that $T=1$, and let $\theta>0$ be fixed. Since, by Lemma 5.1,

$$\mathrm{E}\big(M_k^2\mathbb{1}_{\{|(m_k-1)X_{k-1}|>\theta n\}}\,\big|\,\mathcal{F}_{k-1}\big)\le(\theta n)^{-1}|m_k-1|\left(\sigma_k^2X_{k-1}^2+b_k^2X_{k-1}\right)$$

in order to prove (5.25), by Markov's inequality, it is enough to show that

$$n^{-3}\sum_{k=1}^{n}|m_k-1|\left(\sigma_k^2\mathrm{E}\left(X_{k-1}^2\right)+b_k^2\mathrm{E}(X_{k-1})\right)\le Cn^{-1}\sum_{k=1}^{n}|m_k-1|\left(\sigma_k^2+k^{-1}b_k^2\right)\to 0$$

as $n\to\infty$, where we applied (i) and (ii) of Proposition 5.1. This follows from the Toeplitz theorem since $\{n^{-1}\sigma_k^2 : k=1,\dots,n, n\in\mathbb{N}\}$ and $\{(nk)^{-1}b_k^2 : k=1,\dots,n, n\in\mathbb{N}\}$ are Toeplitz kernels by assumption B and $|m_k-1|\to 0$ as $k\to\infty$ by assumption A. Similarly, in order to prove (5.26) it is enough to show that

$$n^{-3}\sum_{k=1}^{n}\lambda_k\left(\sigma_k^2\mathrm{E}(X_{k-1})+b_k^2\right)\le Cn^{-2}\sum_{k=1}^{n}\lambda_k\left(\sigma_k^2+k^{-1}b_k^2\right)\to 0$$

as $n\to\infty$ which follows from assumptions A and B by Lemma 5.4. Finally, to prove (5.27) we apply the decomposition (5.9) and the inequality (5.24). Thus, it is enough to show that

$$L_n^{i,j}:=n^{-2}\sum_{k=1}^{n}\mathrm{E}\big((\zeta_k^i)^2\mathbb{1}_{\{|\zeta_k^j|>\theta n\}}\,\big|\,\mathcal{F}_{k-1}\big)\xrightarrow{\mathrm{P}}0$$

as $n\to\infty$ for $i,j=1,2$, where ζ_k^i is defined by (5.9). Define the random variables $\zeta_{k,j}^1:=\zeta_k^1-(\xi_{k,j}-m_k)$, $j=1,\dots,X_{k-1}$, $k\in\mathbb{N}$. In the case of $i=j=1$ we have $\mathrm{E}(L_n^{1,1})\le F_n+G_n+H_n$, $n\in\mathbb{N}$, where

$$F_n:=n^{-2}\sum_{k=1}^{n}\mathrm{E}\left(\sum_{j=1}^{X_{k-1}}|\xi_{k,j}-m_k|^2\mathbb{1}_{\{|\xi_{k,j}-m_k|>\theta n/2\}}\right),$$

$$G_n:=n^{-2}\sum_{k=1}^{n}\mathrm{E}\left(\sum_{j=1}^{X_{k-1}}|\xi_{k,j}-m_k|^2\mathbb{1}_{\{|\zeta_{k,j}^1|>\theta n/2\}}\right),$$

$$H_n := 2n^{-2}\sum_{k=1}^{n}\mathrm{E}\left(|S_k|\mathbb{1}_{\{|\zeta_k^1|>\theta n\}}\right), \quad S_k := \sum_{i=2}^{X_{k-1}}\sum_{j=1}^{i-1}(\xi_{k,i}-m_k)(\xi_{k,j}-m_k)\ .$$

Since $\{\xi_{k,j} : j \in \mathbb{N}\}$ are identically distributed for each $k \in \mathbb{N}$, by (i) of Proposition 5.1, we have

$$F_n \le Cn^{-1}\sum_{k=1}^{n}\mathrm{E}\left(|\xi_{k,1}-m_k|^2\mathbb{1}_{\{|\xi_{k,1}-m_k|>\theta n/2\}}\right)\ , \tag{5.28}$$

where the right hand side tends to 0 since the Lindeberg condition holds for $\{n^{-1}\xi_{k,j} : k,j = 1,\dots,n, n \in \mathbb{N}\}$. Since $\xi_{k,j}-m_k$ and $\zeta_{k,j}^1$ are independent for all $k,j \in \mathbb{N}$, by Proposition 5.1, assumption B and Lemma 5.4 we have

$$\begin{aligned} G_n &\le n^{-2}\sum_{k=1}^{n}\mathrm{E}\left(\sum_{j=1}^{X_{k-1}}|\xi_{k,j}-m_k|^2 4(\theta n)^{-2}(\zeta_{k,j}^1)^2\right) \\ &= 4\theta^{-2}n^{-4}\sum_{k=1}^{n}\sigma_k^4\mathrm{E}(X_{k-1}(X_{k-1}-1)) \le Cn^{-2}\sum_{k=1}^{n}\sigma_k^4 \to 0 \end{aligned}$$

as $n \to \infty$. Since $\mathrm{E}(S_k^2) = 2^{-1}\sigma_k^4\mathrm{E}(X_{k-1}(X_{k-1}-1))$ and $\mathrm{E}((\zeta_k^1)^2) = \sigma_k^2\mathrm{E}(X_{k-1})$, $k \in \mathbb{N}$, by the Cauchy-Schwarz and Markov inequalities and Proposition 5.1, we have

$$\begin{aligned} H_n &\le 2n^{-2}\sum_{k=1}^{n}\left(\mathrm{E}(S_k^2)\mathsf{P}(|\zeta_k^1|>\theta n)\right)^{1/2} \le Cn^{-3}\sum_{k=1}^{n}\left(\mathrm{E}(S_k^2)\mathrm{E}(\zeta_k^1)^2\right)^{1/2} \\ &= Cn^{-3}\sum_{k=1}^{n}\sigma_k^3\left(\mathrm{E}(X_{k-1}(X_{k-1}-1))\mathrm{E}(X_{k-1})\right)^{1/2} \le Cn^{-3/2}\left(\sum_{k=1}^{n}\sigma_k^4\sum_{k=1}^{n}\sigma_k^2\right)^{1/2}. \end{aligned}$$

Here, the right hand side tends to 0 by assumption B and Lemma 5.4. In the case of $i=1, j=2$ we have by the Markov inequality that

$$\mathrm{E}\left((\zeta_k^1)^2\mathbb{1}_{\{|\zeta_k^2|>\theta n\}}\middle|\mathcal{F}_{k-1}\right) = \sigma_k^2X_{k-1}\mathsf{P}\left(|\varepsilon_k-\lambda_k|>\theta n\right) \le (\theta n)^{-2}\sigma_k^2X_{k-1}b_k^2\ .$$

Thus, to prove $L_n^{1,2} \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$, by Markov's inequality and (i) of Proposition 5.1, it is enough to see that

$$n^{-4}\sum_{k=1}^{n}\sigma_k^2b_k^2\mathrm{E}(X_{k-1}) \le Cn^{-2}\sum_{k=1}^{n}\sigma_k^2\left(k^{-1}b_k^2\right) \to 0 \tag{5.29}$$

as $n \to \infty$ which follows from assumption B by Lemma 5.4. In the case of $i = 2$, $j = 1$ we have

$$\mathrm{E}\Big((\zeta_k^2)^2 \mathbb{1}_{\{|\zeta_k^1| > \theta n\}} \big| \mathcal{F}_{k-1}\Big) \le (\theta n)^{-2} \mathrm{E}\left((\zeta_k^1 \zeta_k^2)^2 \big| \mathcal{F}_{k-1}\right) = (\theta n)^{-2} \sigma_k^2 X_{k-1} b_k^2 \,.$$

Hence, by (5.29), we have $\mathrm{E}(L_n^{2,1}) \to 0$ which implies $L_n^{2,1} \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. Finally, in the case of $i = j = 2$ we have

$$L_n^{2,2} = n^{-2} \sum_{k=1}^{n} \mathrm{E}\left(|\varepsilon_k - \lambda_k|^2 \mathbb{1}_{\{|\varepsilon_k - \lambda_k| > \theta n\}}\right) \to 0$$

as $n \to \infty$ because the Lindeberg condition holds for $\{n^{-1}\varepsilon_k : k = 1, \ldots, n, n \in \mathbb{N}\}$. Thus, we finished the proof. □

Proof of Theorem 5.2 We apply Theorem 5.3 (see Appendix) with the choice $U_k^n := n^{-1/2}(\widetilde{X}_k - \widetilde{X}_{k-1})$, $k, n \in \mathbb{N}$, $\mathcal{F}_k^n := \mathcal{F}_k$, $k \in \mathbb{Z}_+, n \in \mathbb{N}$, and $\beta(t,x) := 0$, $\gamma(t,x) := \tilde{\sigma}$, $t \in \mathbb{R}_+, x \in \mathbb{R}$. (Clearly, $(U_k^n)_{k\in\mathbb{N}}$ is not a sequence of martingale differences, hence the standard martingale central limit theorem cannot be applied immediately.) Then $\mathcal{U}^n = \widetilde{\mathcal{X}}^n$, $n \in \mathbb{N}$, and $\mathcal{U} = \tilde{\sigma}\mathcal{W}$, where $(\mathcal{W}_t)_{t\in\mathbb{R}_+}$ is a standard Wiener process. We have to prove for all $T > 0$ that

$$\sup_{t\in[0,T]} \left| \sum_{k=1}^{\lfloor nt \rfloor} \mathrm{E}\Big(U_k^n \,|\, \mathcal{F}_{k-1}\Big) \right| \xrightarrow{\mathrm{P}} 0 \,, \tag{5.30}$$

$$\sup_{t\in[0,T]} \left| \sum_{k=1}^{\lfloor nt \rfloor} \mathsf{Var}\Big(U_k^n \,|\, \mathcal{F}_{k-1}\Big) - \tilde{\sigma}^2 t \right| \xrightarrow{\mathrm{P}} 0 \tag{5.31}$$

as $n \to \infty$ and (5.40) holds for the pair $(U_k^n)_{k\in\mathbb{N}}$, $(\mathcal{F}_k)_{k\in\mathbb{Z}_+}$, $n \in \mathbb{N}$.

We have $\mathrm{E}(U_k^n \,|\, \mathcal{F}_{k-1}) = n^{-1/2}(m_k - 1)\widetilde{X}_{k-1}$, $k, n \in \mathbb{N}$. In order to prove (5.30) we can suppose that $T = 1$ and, by Markov's inequality, it is enough to show that

$$n^{-1/2} \sum_{k=1}^{n} |m_k - 1| \mathrm{E}|\widetilde{X}_{k-1}| \le C n^{-1/2} \sum_{k=1}^{n} k^{1/2} |m_k - 1| \to 0$$

as $n \to \infty$ which follows by Kronecker's lemma. Note that we applied the Lyapunov inequality $\mathrm{E}|\widetilde{X}_n| \le (\mathsf{Var}(X_n))^{1/2} = O(n^{1/2})$ together with (iii) of Proposition 5.1.

By Lemma 5.1 we obtain $\mathsf{Var}(U_k^n \,|\, \mathcal{F}_{k-1}) = n^{-1}\mathrm{E}(M_k^2 \,|\, \mathcal{F}_{k-1}) = n^{-1}(\sigma_k^2 X_{k-1} + b_k^2)$, $k, n \in \mathbb{N}$. Thus, in order to prove (5.31) it is enough to show that

$$\sup_{t\in[0,T]} \left| n^{-1} \sum_{k=1}^{\lfloor nt \rfloor} \sigma_k^2 \widetilde{X}_{k-1} \right| \xrightarrow{\mathrm{P}} 0 \,, \tag{5.32}$$

$$\sup_{t\in[0,T]}\left|n^{-1}\sum_{k=1}^{\lfloor nt\rfloor}\sigma_k^2\mathrm{E}(X_{k-1})-\sigma^2\lambda t\right|\to 0\,, \tag{5.33}$$

$$\sup_{t\in[0,T]}\left|n^{-1}\sum_{k=1}^{\lfloor nt\rfloor}b_k^2-b^2t\right|\to 0 \tag{5.34}$$

as $n\to\infty$. To prove (5.32) we can also suppose that $T=1$ and, by Lyapunov's and Markov's inequalities together with (iii) of Proposition 5.1, it is enough to show that

$$n^{-1}\sum_{k=1}^{n}\sigma_k^2\mathrm{E}|\widetilde{X}_{k-1}|\le n^{-1}\sum_{k=1}^{n}\sigma_k^2\left(\mathsf{Var}(X_{k-1})\right)^{1/2}\le Cn^{-1}\sum_{k=1}^{n}\sigma_k^2k^{1/2}\to 0$$

as $n\to\infty$. This follows from the Toeplitz theorem since $\{n^{-1}k\sigma_k^2 : k=1,\dots,n,n\in\mathbb{N}\}$ is a Toeplitz kernel by assumption C. Since $n^{-1}\sum_{k=1}^{n}\sigma_k^2\mathrm{E}(X_{k-1})\to\sigma^2\lambda$ as $n\to\infty$ by assumption C, (i) of Proposition 5.1, and the Toeplitz theorem, Lemma 5.3 implies (5.33). Finally, Lemma 5.3 also implies (5.34) by assumption C.

The conditional Lindeberg condition (5.40) for the pair $(U_k^n)_{k\in\mathbb{N}}, (\mathcal{F}_k)_{k\in\mathbb{Z}_+}$, $n\in\mathbb{N}$, can be proved similarly as in the proof of Theorem 5.1. Since $\widetilde{X}_k-\widetilde{X}_{k-1}=(m_k-1)\widetilde{X}_{k-1}+M_k$, $k\in\mathbb{N}$, it is enough to prove, for all $\theta>0$, that

$$n^{-1}\sum_{k=1}^{\lfloor nT\rfloor}(m_k-1)^2\widetilde{X}_{k-1}^2\xrightarrow{\mathrm{P}}0\,, \tag{5.35}$$

$$n^{-1}\sum_{k=1}^{\lfloor nT\rfloor}\mathrm{E}\big(M_k^2\mathbb{1}_{\{|(m_k-1)\widetilde{X}_{k-1}|>\theta n^{1/2}\}}\,\big|\,\mathcal{F}_{k-1}\big)\xrightarrow{\mathrm{P}}0\,, \tag{5.36}$$

$$n^{-1}\sum_{k=1}^{\lfloor nT\rfloor}\mathrm{E}\big(M_k^2\mathbb{1}_{\{|M_k|>\theta n^{1/2}\}}\,\big|\,\mathcal{F}_{k-1}\big)\xrightarrow{\mathrm{P}}0 \tag{5.37}$$

as $n\to\infty$. Clearly, it is again enough to consider the case when $T=1$. Using (iii) of Proposition 5.1 and $\sum_{k=1}^{\infty}(m_k-1)^2<+\infty$, the Kronecker lemma implies

$$n^{-1}\sum_{k=1}^{n}(m_k-1)^2\mathrm{E}\big(\widetilde{X}_{k-1}^2\big)\le Cn^{-1}\sum_{k=1}^{n}k(m_k-1)^2\to 0$$

as $n\to\infty$. Thus, by Markov's inequality, we obtain (5.35). By Lemma 5.1 we have

$$\mathrm{E}\big(M_k^2\mathbb{1}_{\{|(m_k-1)\widetilde{X}_{k-1}|>\theta n^{1/2}\}}\,\big|\,\mathcal{F}_{k-1}\big)\le\theta^{-1}n^{-1/2}|m_k-1||\widetilde{X}_{k-1}|(\sigma_k^2X_{k-1}+b_k^2)\,.$$

Moreover, $\mathrm{E}(|\widetilde{X}_n|X_n) \leq (\mathsf{Var}(X_n)\mathrm{E}(X_n^2))^{1/2} = O(n^{3/2})$ and $\mathrm{E}|\widetilde{X}_n| \leq (\mathsf{Var}(X_n))^{1/2} = O(n^{1/2})$ by the Cauchy-Schwarz and Lyapunov inequalities and Proposition 5.1. Thus, again by Markov's inequality, to prove (5.36) it is enough to show that

$$n^{-3/2}\sum_{k=1}^{n}\sigma_k^2|m_k-1|\mathrm{E}(|\widetilde{X}_{k-1}|X_{k-1}) \leq Cn^{-1}\sum_{k=1}^{n}k\sigma_k^2|m_k-1| \to 0\,,$$

$$n^{-3/2}\sum_{k=1}^{n}b_k^2|m_k-1|\mathrm{E}|\widetilde{X}_{k-1}| \leq Cn^{-1}\sum_{k=1}^{n}b_k^2|m_k-1| \to 0$$

as $n \to \infty$ which follows from the Toeplitz theorem since $\{n^{-1}k\sigma_k^2 : k = 1,\dots,n, n \in \mathbb{N}\}$ and $\{n^{-1}b_k^2 : k = 1,\dots,n, n \in \mathbb{N}\}$ are Toeplitz kernels by assumption C and $|m_k - 1| \to 0$ as $k \to \infty$ by assumption A. Finally, we prove (5.37) similarly to the proof of (5.27). For each $n \in \mathbb{N}$, define $\widetilde{L}_n^{i,j}$, $i,j = 1,2$, and $\widetilde{F}_n, \widetilde{G}_n, \widetilde{H}_n$ similarly to $L_n^{i,j}$, $i,j = 1,2$, and F_n, G_n, H_n, by replacing the normalizing factor n^{-2} with n^{-1} and n with $n^{1/2}$ in the indicator functions, respectively. It remains to prove $\widetilde{L}_n^{i,j} \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$ for $i,j = 1,2$. Similarly to (5.28), we have

$$\widetilde{F}_n \leq C\sum_{k=1}^{n}\mathrm{E}\left(|\xi_{k,1}-m_k|^2\mathbb{1}_{\{|\xi_{k,1}-m_k|>\theta n^{1/2}/2\}}\right) \to 0$$

as $n \to \infty$ since the Lindeberg condition holds for $\{n^{-1/2}\xi_{k,j} : k,j = 1,\dots,n, n \in \mathbb{N}\}$. Moreover, we have

$$\widetilde{G}_n \leq 4\theta^{-2}n^{-2}\sum_{k=1}^{n}\sigma_k^4\mathrm{E}(X_{k-1}(X_{k-1}-1)) \leq Cn^{-2}\sum_{k=1}^{n}\left(k\sigma_k^2\right)^2 \to 0$$

as $n \to \infty$ by Lemma 5.4 and assumption C. By the Cauchy-Schwarz inequality we have

$$\widetilde{H}_n \leq Cn^{-3/2}\sum_{k=1}^{n}\sigma_k^3k^{3/2} \leq Cn^{-3/2}\left(\sum_{k=1}^{n}k\sigma_k^2\sum_{k=1}^{n}\left(k\sigma_k^2\right)^2\right)^{1/2},$$

where the right hand side tends to 0 as $n \to \infty$ by assumption C and Lemma 5.4. Thus, we proved $\widetilde{L}_n^{1,1} \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$. Since

$$\mathrm{E}\big((\zeta_k^1)^2\mathbb{1}_{\{|\zeta_k^2|>\theta n^{1/2}\}}\big|\mathcal{F}_{k-1}\big) = \sigma_k^2X_{k-1}\mathsf{P}\big(|\varepsilon_k-\lambda_k| > \theta n^{1/2}\big) \leq \theta^{-2}n^{-1}\sigma_k^2X_{k-1}b_k^2$$

in order to prove the case $i = 1, j = 2$ it is enough to show that

$$n^{-2} \sum_{k=1}^{n} \sigma_k^2 b_k^2 \mathrm{E}(X_{k-1}) \leq C n^{-2} \sum_{k=1}^{n} (k\sigma_k^2) b_k^2 \to 0 \tag{5.38}$$

as $n \to \infty$ which follows from assumption C by Lemma 5.4. In the case of $i = 2$, $j = 1$ we have

$$\mathrm{E}\big((\zeta_k^2)^2 \mathbb{1}_{\{|\zeta_k^1| > \theta n^{1/2}\}} \big| \mathcal{F}_{k-1}\big) \leq \theta^{-2} n^{-1} \mathrm{E}\left((\zeta_k^1 \zeta_k^2)^2 \big| \mathcal{F}_{k-1}\right) = \theta^{-2} n^{-1} \sigma_k^2 X_{k-1} b_k^2 \, .$$

Thus, $\widetilde{L}_n^{2,1} \xrightarrow{\mathrm{P}} 0$ as $n \to \infty$ follows from (5.38) by Markov's inequality. Finally,

$$\widetilde{L}_n^{2,2} = n^{-1} \sum_{k=1}^{n} \mathrm{E}\big(|\varepsilon_k - \lambda_k|^2 \mathbb{1}_{\{|\varepsilon_k - \lambda_k| > \theta n^{1/2}\}}\big) \to 0$$

as $n \to \infty$ since the Lindeberg condition holds for $\{n^{-1/2}\varepsilon_k : k = 1, \ldots, n, n \in \mathbb{N}\}$.
□

Acknowledgements The author would like to thank the anonymous reviewer for drawing his attention to the close connection between branching processes with immigration and controlled branching processes and to the papers González and del Puerto [3] and González et al. [4].

This paper was finished when the author was a researcher at the Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences.

Appendix

In the proofs we need the following result about convergence of random step processes towards a diffusion process, see Ispány and Pap [6, Corollary 2.2].

Theorem 5.3 *Let* $\beta : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ *and* $\gamma : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ *be continuous functions. Assume that uniqueness in the sense of probability law holds for the SDE*

$$\mathrm{d}\mathcal{U}_t = \beta(t, \mathcal{U}_t)\,\mathrm{d}t + \gamma(t, \mathcal{U}_t)\,\mathrm{d}\mathcal{W}_t \, , \qquad t \in \mathbb{R}_+ \, , \tag{5.39}$$

with initial value $\mathcal{U}_0 = u_0$ *for all* $u_0 \in \mathbb{R}$, *where* $(\mathcal{W}_t)_{t \in \mathbb{R}_+}$ *is a standard Wiener process. Let* $(\mathcal{U}_t)_{t \in \mathbb{R}_+}$ *be a solution of* (5.39) *with initial value* $\mathcal{U}_0 = 0$.

For each $n \in \mathbb{N}$, *let* $(U_k^n)_{k \in \mathbb{N}}$ *be a sequence of random variables adapted to a filtration* $(\mathcal{F}_k^n)_{k \in \mathbb{Z}_+}$. *Let* $\mathcal{U}_t^n := \sum_{k=1}^{\lfloor nt \rfloor} U_k^n$, $t \in \mathbb{R}_+, n \in \mathbb{N}$. *Suppose* $\mathrm{E}\big(|U_k^n|^2\big) < \infty$

for all $n, k \in \mathbb{N}$. *Suppose that, for each* $T > 0$,

$$\sup_{t\in[0,T]} \left| \sum_{k=1}^{\lfloor nt \rfloor} \mathrm{E}\big(U_k^n \,|\, \mathcal{F}_{k-1}^n\big) - \int_0^t \beta(s, \mathcal{U}_s^n)\,\mathrm{d}s \right| \xrightarrow{\mathrm{P}} 0 \,,$$

$$\sup_{t\in[0,T]} \left| \sum_{k=1}^{\lfloor nt \rfloor} \mathsf{Var}\big(U_k^n \,|\, \mathcal{F}_{k-1}^n\big) - \int_0^t \big(\gamma(s, \mathcal{U}_s^n)\big)^2\,\mathrm{d}s \right| \xrightarrow{\mathrm{P}} 0 \,,$$

$$\sum_{k=1}^{\lfloor nT \rfloor} \mathrm{E}\big(|U_k^n|^2 \mathbb{1}_{\{|U_k^n|>\theta\}} \,\big|\, \mathcal{F}_{k-1}^n\big) \xrightarrow{\mathrm{P}} 0 \quad \textit{for all } \theta > 0 \,. \tag{5.40}$$

Then $\mathcal{U}^n \xrightarrow{\mathcal{L}} \mathcal{U}$ *as* $n \to \infty$.

References

1. Billingsley, P.: Convergence of Probability Measures, 2nd edn. Wiley, New York (1999)
2. Fearn, D.H.: Galton-Watson processes with generation dependence. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 159–172. University of California Press, Berkeley (1972)
3. González, M., del Puerto, I.: Diffusion approximation of an array of controlled branching processes. Methodol. Comput. Appl. Probab. **14**, 843–861 (2012)
4. González, M., Minuesa, C., Mota, M., del Puerto, I., Ramos, A.: An inhomogeneous controlled branching process. Lith. Math. J. **55**(1), 61–71 (2015)
5. Haccou, P., Jagers, P., Vatutin, V.A.: Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press, Cambridge (2005)
6. Ispány, M., Pap, G.: A note on weak convergence of step processes. Acta Math. Hung. **126**(4), 381–395 (2010)
7. Ispány, M., Pap, G., van Zuijlen, M.C.A.: Fluctuation limit of branching processes with immigration and estimation of the means. Adv. Appl. Probab. **37**(2), 523–538 (2005)
8. Jagers, P.: Galton-Watson processes in varying environments. J. Appl. Probab. **11**(1), 174–178 (1974)
9. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus, 2nd edn. Springer, Berlin (1991)
10. Keiding, N., Nielsen, J.: Branching processes with varying and random geometric offspring distributions. J. Appl. Probab. **12**(1), 135–141 (1975)
11. Mitov, K.V., Omey, E.: A branching process with immigration in varying environments. Commun. Stat. Theory Methods **43**(24), 5211–5225 (2014)
12. Rahimov, I.: Random Sums and Branching Stochastic Processes. Lecture Notes in Statistics. Springer, New York (1995)
13. Rahimov, I.: Functional limit theorems for critical processes with immigration. Adv. Appl. Probab. **39**(4), 1054–1069 (2007)
14. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion, 3rd edn., corrected 2nd printing. Springer, Berlin (2001)
15. Wei, C.Z., Winnicki, J.: Some asymptotic results for the branching process with immigration. Stoch. Process. Appl. **31**(2), 261–282 (1989)

Chapter 6
Subcritical Branching Processes in Random Environment

Vladimir Vatutin

Mathematics Subject Classification (2000): 60J80, 60F17, 60K37

6.1 Introduction

For the first time the model of a branching process in random environment (BPRE) was considered by Smith and Wilkinson [20] in 1969 who dealt with the case of i.i.d. random environments. Then Athreya and Karlin [10, 11] in 1971 have studied BPRE's evolving in a more general environment. Since then a great number of articles have been published investigating the behavior of branching processes in random environment (see, for instance, surveys [13, 24, 25]). The present survey, having certain intersections with [25] is devoted to a description of the most recent results for subcritical branching processes in random environment obtained mainly by the author in collaboration with other scientists.

There are two possibilities to study stochastic processes in random environment: the quenched and annealed approaches. Under the *quenched approach* characteristics of a BPRE such as the survival probability at moment n or the distribution of the number of particles at this moment are treated as random variables or random measures where the source of randomness is due to uncertainty in possible realizations of the environment. Under the *annealed approach* the aim is to investigate the mean values of the mentioned characteristics using the averaging over possible realizations of the environment.

Now we give the formal definition of the BPRE's.

V. Vatutin (✉)
Steklov Mathematical Institute, Gubkin Street 8, 119991 Moscow, Russia
e-mail: vatutin@mi.ras.ru

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_6

Let $\mathfrak{N}$ be the space of probability measures on $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Equipped with the metric of total variation $\mathfrak{N}$ becomes a Polish space. Let π be a random variable taking values in $\mathfrak{N}$. An infinite sequence $\Pi = (\pi_1, \pi_2, \ldots)$ of i.i.d. copies of π is said to form a *random environment*. A sequence of $\mathbb{N}_0$-valued random variables $Z_0, Z_1, \ldots$ is called a *branching process in the random environment* Π, if Z_0 is independent of Π and, given Π, the process $\mathbf{Z} = (Z_0, Z_1, \ldots)$ is a Markov chain with

$$\mathcal{L}\left(Z_n \mid Z_{n-1} = z_{n-1}, \Pi = (\pi_1, \pi_2, \ldots)\right) = \mathcal{L}\left(\xi_{n1} + \cdots + \xi_{nz_{n-1}}\right) \tag{6.1}$$

for every $n \geq 1$, $z_{n-1} \in \mathbb{N}_0$ and $\pi_1, \pi_2, \ldots \in \mathfrak{N}$, where $\xi_{n1}, \xi_{n2}, \ldots$ are i.i.d. random variables with distribution π_n. Thus,

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{ni} \tag{6.2}$$

and, given the environment, $\mathbf{Z}$ is an ordinary inhomogeneous Galton-Watson process. We will denote the corresponding probability measure and expectation on the underlying probability space by $\mathbb{P}$ and $\mathbb{E}$, respectively.

Let

$$f(s) = f(\pi; s) = \sum_{k=0}^{\infty} \pi(\{k\}) s^k,$$

$$f_n(s) = f_n(\pi_n; s) = \sum_{k=0}^{\infty} \pi_n(\{k\}) s^k, n = 1, 2, \ldots$$

be a tuple of *random* probability generating functions. Denote

$$X = \log f'(1), \qquad X_n = \log f_n'(1),\ n = 1, 2, \ldots,$$

and introduce the so-called *associated random walk* (ARW) $\mathbf{S} = \{S_n, n = 0, 1, \ldots\}$ specified by the formulas

$$S_0 = 0,\ S_n = X_1 + \cdots + X_n,\ n \geq 1.$$

Throughout we will denote the probability and expectation conditioned on the environment by

$$\mathcal{P}(\cdot) = \mathbb{P}\left(\cdot \mid f_1, f_2, \ldots\right), \quad \mathcal{E}[\cdot] = \mathbb{E}\left[\cdot \mid f_1, f_2, \ldots\right].$$

Clearly,

$$\mathcal{E}\left[Z_n\right] = e^{S_n}.$$

This representation leads to the following classification of branching processes in random environment: a BPRE is called

1. supercritical if $\lim_{n\to\infty} S_n = +\infty$ with probability 1;
2. subcritical if $\lim_{n\to\infty} S_n = -\infty$ with probability 1;
3. non-degenerate critical if $\limsup_{n\to\infty} S_n = +\infty$ and $\liminf_{n\to\infty} S_n = -\infty$ both with probability 1;
4. degenerate critical if $S_n = 0$ for all n.

This classification, due to [6], is based on the crucial fact that *any* random walk with i.i.d. increments can be attributed to one of these four classes. Since

$$\mathcal{P}(Z_n > 0) = \min_{1\le k\le n} \mathcal{P}(Z_k > 0) \le \min_{1\le k\le n} \mathcal{E}[Z_k] = e^{\min(S_1,\dots,S_n)}$$

it follows that the extinction probability of subcritical and non-degenerate critical BPRE's equals 1 a.s.

In many situations we may express the classification above in terms of the moment generating function

$$\varphi(t) = \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\left(f'(\pi;1)\right)^t\right], \quad t \in (-\infty, +\infty).$$

Clearly, $\varphi'(0) = \mathbb{E}[X]$ if both of the quantities exist. If this is the case we say that a BPRE is

- supercritical if $\varphi'(0) = \mathbb{E}[X] = \mathbb{E}[\log f'(1)] > 0$,
- critical if $\varphi'(0) = \mathbb{E}[X] = \mathbb{E}[\log f'(1)] = 0$,
- subcritical if $\varphi'(0) = \mathbb{E}[X] = \mathbb{E}[\log f'(1)] = -a < 0$.

Subcritical BPRE's admit an additional classification which is based on the properties of $\varphi(t)$. Namely, let

$$\rho_+ = \sup\{t \ge 0 : \varphi(t) < \infty\} \text{ and } \rho_- = \inf\{t : \varphi(t) < \infty\}$$

and let β be the point where $\varphi(t)$ attains its minimal value on the interval

$$[0, \rho_+ \wedge 1] \subseteq [\rho_-, \rho_+].$$

Then a subcritical BPRE is called

- weakly subcritical if $\beta \in (0, \rho_+ \wedge 1)$,
- intermediately subcritical if $\beta = \rho_+ \wedge 1 > 0$ and $\varphi'(\beta-) = 0$,
- strongly subcritical if $\beta = \rho_+ \wedge 1 > \rho_-$ and $\varphi'(\beta-) < 0$,
- boundary subcritical if $\beta = \rho_- = \rho_+ = 0$.

For these subcritical BPRE's we give a survey of some recent results concerning the following problems **under the annealed approach**:

- the asymptotic behavior of the survival probabilities $\mathbb{P}(Z_n > 0)$ as $n \to \infty$;
- the form of the environments providing the survival, i.e. we list a number of theorems for the conditional distributions $\mathbb{P}(S_m \in dx|Z_n > 0)$, as n and m tend to infinity in an appropriate way;
- conditional limit theorems of the form $\mathbb{P}(Z_n \in dx\,|Z_n > 0)$ as $n \to \infty$;
- functional conditional limit theorems of the form $\mathbb{P}(Z_{nt} \in dx, 0 \leq t \leq 1|Z_n > 0)$ as $n \to \infty$;
- conditional limit theorems of the form $\mathbb{P}(T_x \in dy\,|T_x < \infty)$ as $x \to \infty$, where $T_x = \min\{k \geq 1 : Z_k > x\}$.

Setting

$$\gamma = \mathbb{E}[e^{\beta X}] \tag{6.3}$$

and observing that $\mathbb{E}\left[e^{\beta S_n}\right] = \gamma^n$ we introduce the following auxiliary measure $\mathbf{P}$ with expectation $\mathbf{E}$. For any $n \in \mathbb{N}$ and any measurable, bounded function $\psi : \Delta^n \times \mathbb{N}_0^{n+1} \to \mathbb{R}$, the measure $\mathbf{P}$ is given by

$$\mathbf{E}[\psi(\pi_1,\dots,\pi_n;Z_0,\dots,Z_n)] = \gamma^{-n}\mathbb{E}\big[\psi(\pi_1,\dots,\pi_n;Z_0,\dots,Z_n)e^{\beta S_n}\big].$$

Notice that if $\mathbf{E}[X] = 0$ then $\mathbf{S}$ is a recurrent random walk under $\mathbf{P}$.

Condition 6.1 *The distribution of X with respect to $\mathbf{P}$ is non-lattice, has zero mean and belongs to the domain of attraction of a stable law with index $\alpha \in (1,2]$.*

Under Condition 6.1 there exists an increasing sequence of positive numbers a_n regularly varying at infinity

$$a_n = n^{1/\alpha} l(n) \tag{6.4}$$

such that the scaled ARW $a_n^{-1} S_{nt}$ $\mathbf{P}$-weakly converges to a strictly stable Lévy process $\mathbf{L} = (L(t),\ t \geq 0)$ with parameter α. Here and elsewhere in the expressions like S_{nt} the index nt is understood as its integer part.

Denote

$$L_n = \min(S_1,\dots,S_n)\,.$$

One of the main ideas in finding the asymptotic behavior of the survival probability of subcritical BPRE's within the framework of the annealed approach is to show that under natural conditions

$$\mathbb{P}(Z_n > 0) \sim \theta\,\mathbb{P}(L_n \geq 0)\,, \quad n \to \infty, \tag{6.5}$$

where $\theta > 0$. This is done by splitting trajectories of the ARW at the point of global minimum on the interval $[0, n]$. For the annealed approach the trajectory splitting method was first used for the critical BPRE's by Dyakonova et al. [16] and then further developed by Afanasyev et al. [6].

6.1.1 Weakly Subcritical Case

Let $\xi = \xi(\pi)$ be the offspring number of a particle evolving in the environment π. By definition

$$\mathbb{E}[s^{\xi}] = \mathbb{E}[f(s)].$$

To state the known results for the weakly subcritical case we need to impose a higher moment assumption on the environment in terms of the standardized truncated second moment of ξ

$$\zeta(a) \;=\; \frac{\mathbb{E}\left[\xi^2 1_{\{\xi \geq a\}}\right]}{(\mathbb{E}\,[\xi])^2}. \tag{6.6}$$

Condition 6.2 *For some positive ε and a*

$$\mathbf{E}[(\ln^+ \zeta(a))^{\alpha+\varepsilon}] \;<\; \infty.$$

It was proved by Afanasyev et al. [8] for the weakly subcritical BPRE's that under Conditions 6.1 and 6.2 there exist numbers $0 < \theta, \theta' < \infty$ such that

$$\mathbb{P}\,(Z_n > 0) \;\sim\; \theta\,\mathbb{P}\,(L_n \geq 0) \sim\; \theta' \frac{\gamma^n}{n a_n}\,,$$

where γ is from (6.3) and a_n are from (6.4). In addition, it was established under the same assumptions that the conditional laws

$$\mathcal{L}(Z_n \,|\, Z_n > 0), n \geq 1,$$

converge weakly to some probability law with the support on the set of positive integers and the sequence $\mathbb{E}[Z_n^{\vartheta} \,|\, Z_n > 0]$ is bounded for any $\vartheta < \beta$. The last implies convergence of the prelimiting moments of orders $\vartheta < \beta$ to the corresponding moments of the limit distribution.

For integers $0 \leq r \leq n$ let

$$X_t^{r,n} \;=\; Z_{r+(n-r)t} \cdot e^{-S_{r+(n-r)t}}\,, \quad t \in [0,1]\,, \tag{6.7}$$

be the rescaled generation size process. It was shown in [8] that under Conditions 6.1 and 6.2, as $n \to \infty$

$$\mathcal{L}\big(X_t^{r_n,n},\ t \in [0,1] \,\big|\, Z_n > 0\big) \Longrightarrow \mathcal{L}\big(W_t,\ t \in [0,1]\big),$$

where the symbol $\Longrightarrow$ means weak convergence in the space $D[0,1]$ endowed with Skorokhod topology, and $r_1, r_2, \ldots$ are natural numbers such that $r_n < n/2$ and $r_n \to \infty$. Here $W_t = W$ a.s. for all $t \in [0,1]$ and $\mathbb{P}(0 < W < \infty) = 1$. Earlier versions of these results can be found in [1, 17].

6.1.2 Intermediately Subcritical Case

Afanasyev et al. [9] proved that given $\beta = 1$ and Conditions 6.1, 6.2 there are a constant $0 < \theta < \infty$ and a sequence $l(n)$, $n = 1, 2, \ldots$ slowly varying at infinity such that

$$\mathbb{P}(Z_n > 0) \sim \theta\gamma^n \mathbf{P}(S_n < \min(S_1, \ldots, S_{n-1})) \sim \frac{\gamma^n l(n)}{n^{1-1/\alpha}}, \quad n \to \infty,$$

and that the unscaled population size Z_n conditioned on $\{Z_n > 0\}$ converges in distribution, as $n \to \infty$ to a proper random variable.

These results were first obtained by Vatutin in [21] under slightly different assumptions.

Consider the α-stable Lévy process $\mathbf{L}$ on the interval $0 \le t \le 1$ and let $\mathbf{L}^c = (L^c(t),\ 0 \le t \le 1)$ be the corresponding Lévy process conditioned on having its minimum at time $t = 1$. For the precise definition of such a process, we refer to [9]. Let $e_1, e_2, \ldots$ denote the excursion intervals of $\mathbf{L}^c$ between consecutive local minima and put $j(t) = i$ for $t \in e_i$. It was shown in [9] that Conditions 6.1 and 6.2 imply for $0 < t_1 < t_2 < \ldots < t_k < 1$ and $n \to \infty$:

$$\mathcal{L}\left(\frac{Z_{nt_1}}{\exp\left(S_{nt_1} - \min_{k \le nt_1} S_k\right)}, \ldots, \frac{Z_{nt_k}}{\exp\left(S_{nt_k} - \min_{k \le nt_k} S_k\right)} \,\Big|\, Z_n > 0\right)$$
$$\xrightarrow{d} \mathcal{L}\left(V_{j(t_1)}, \ldots, V_{j(t_k)}\right), \qquad (6.8)$$

where $V_1, V_2, \ldots$ are i.i.d. copes of some strictly positive random variable V. This means that if t_i and t_k belong to one and the same excursion, then $j(t_i)$ and $j(t_k)$ coincide, i.e. $V_{j(t_i)} = V_{j(t_k)}$ with probability 1. Notice that here

$$j(t_1) \le j(t_2) \le \ldots \le j(t_n).$$

For the fractional-linear case this result was earlier obtained by Afanasyev in [2].

Define the process $\mathbf{L}^r = (L^r(t),\ 0 \le t \le 1)$, which is the process $\mathbf{L}^c$ reflected at its current minimum and given by

$$L^r(t) = L^c(t) - \min_{s \le t} L^c(s)\ .$$

The next theorem, established in [14] refines the properties of the trajectories of the intermediately subcritical process $\mathbf{Z}$ given its survival up to a distant moment n:

Theorem 6.3 *Under Conditions 6.1, 6.2, as $n \to \infty$*

$$\mathcal{L}\Big(\frac{\log Z_{nt}}{a_n},\ 0 \le t \le 1 \,\Big|\, Z_n > 0\Big) \Longrightarrow \mathcal{L}(L^r(t),\ 0 \le t \le 1).$$

Let

$$\tau\,(nt) = \min\{i \in [0, nt] :\ S_i = \min\,(0, L_{nt})\}$$

be the left–most point at which the ARW attains its minimal value on the time-interval $[0, nt]$. It was demonstrated in [9] that at the times of consecutive ARW minima the population sizes have discrete limit distributions

$$\mathcal{L}\big(\left(Z_{\tau(nt_1)}, \ldots, Z_{\tau(nt_k)}\right)\ \big|\, Z_n > 0\,\big) \overset{d}{\to} \mathcal{L}\left(Y_{j(t_1)}, \ldots, Y_{j(t_k)}\right),$$

where $Y_1, Y_2, \ldots$ are i.i.d. copies of a random variable Y taking values in $\mathbb{N}$. This, combined with (6.8) shows that the population of the intermediate subcritical process survived up to a distant moment n had passed through a number of bottlenecks when the size of the population dropped down to a small number of individuals while between the bottlenecks the population size has an exponential order (with the parameter of exponentiality proportional to $\sqrt{n}$).

6.1.3 Strongly Subcritical Case

6.1.3.1 The Case $\beta = 1$

Guivarc'h and Liu in [18] have shown for strongly subcritical processes with $\beta = 1$ and satisfying

$$\mathbb{E}\left[\xi \ln^+ \xi\right] < \infty \tag{6.9}$$

that for some $c \in (0, 1]$

$$\mathbb{P}\,(Z_n > 0) \sim c\,(\mathbb{E}\,[\xi])^n\ ,\ n \to \infty. \tag{6.10}$$

This asymptotic formula was originally established by D'Souza and Hambly in [15] under an extra moment assumption. For the fractional-linear case (6.10) was obtained in [1].

Geiger et al. [17] studied the strongly subcritical BPRE's under condition (6.9) and proved that

$$\lim_{n\to\infty} \mathbb{P}\left(Z_n = z \mid Z_n > 0\right) = r_z,\ z \in \mathbb{N},$$

where

$$\sum_{z=1}^{\infty} r_z = 1 \text{ and } m_r = \sum_{z=1}^{\infty} z r_z < \infty.$$

Let

$$\eta = \frac{f''(1)}{(f'(1))^2} = \frac{\mathbb{E}\left[\xi(\xi-1)\right]}{(\mathbb{E}\left[\xi\right])^2}. \tag{6.11}$$

Afanasyev et al. [7] have introduced the following condition in the strongly subcritical case.

Condition 6.4 *Suppose that*

$$\mathbb{E}\left[e^X \ln^+ \eta\right] < \infty.$$

Since $\mathbb{E}\left[e^X\right] < \infty$ for the strongly subcritical case with $\beta = 1$, Condition 6.4 holds, in particular, if the random offspring distribution π has a uniformly bounded support. It also holds if π is a Poisson distribution with random mean, so that $\eta = 1$ a.s., or if π is a geometric distribution on $\mathbb{N}_0$ where $\eta = 2$ a.s.

It was shown in [7] that given

$$0 = i_{n,0} < i_{n,1} < i_{n,2} < \ldots < i_{n,k} < i_{n,k+1} = n$$

the following weak convergence holds

$$\begin{aligned}&\mathcal{L}\big(\left(Z_j\right)_{0\le j\le m}, \left(Z_{i_{n,1}+j}\right)_{0\le j\le m}, \ldots, \left(Z_{i_{n,k}+j}\right)_{0\le j\le m}, \left(Z_{n-j}\right)_{0\le j\le m} \mid Z_n > 0\big) \\ &\xrightarrow{d} \mathcal{L}_{\delta_1}\left(\left(Y_j\right)_{0\le j\le m}\right) \otimes \mathcal{L}_{\hat{r}}\left(\left(Y_j\right)_{0\le j\le m}\right)^{\otimes k} \otimes \mathcal{L}_r\left((\widetilde{Y}_j)_{0\le j\le m}\right),\end{aligned}$$

for every $k, m \in \mathbb{N}_0$ as $\min_{0\le l\le k}\left(i_{n,l+1} - i_{n,l}\right) \to \infty$ and $n \to \infty$. Here, $\mathcal{L}_\mu\left(\left(Y_j\right)_{j\ge 0}\right)$ denotes the law of the Markov chain $\left(Y_j\right)_{j\ge 0}$ with initial distribution μ and transitional probabilities

$$\widehat{P}_{yz} = \frac{z\mathbb{P}\left(\xi_1 + \ldots + \xi_y = z\right)}{y\mathbb{E}\left[\xi\right]}, \quad y, z \in \mathbb{N},$$

where $\xi_1, \ldots, \xi_y$ are independent copies of the offspring number ξ. The stationary distribution of $\left(Y_j\right)_{j\geq 0}$ is

$$\hat{r}_z = \frac{z r_z}{m_r}, \; z \in \mathbb{N}.$$

The process $(\widetilde{Y}_j)_{j\geq 0}$ is the time-reversed Markov chain with transition probabilities

$$\widetilde{P}_{yz} = \frac{\hat{r}_z \widehat{P}_{zy}}{\hat{r}_y}.$$

6.1.3.2 The Case $\beta = 0$

In this part of the survey we consider a subcritical branching process in random environment such that

$$\mathbb{E}\left[Xe^{tX}\right] = \infty$$

for any $t > 0$.

To be more specific, we suppose that $a = -\mathbb{E}[X] > 0$, $\mathbb{E}\left[X^2\right] < \infty$ and, in addition, as $x \to \infty$,

$$A(x) = \mathbb{P}\left(X > x\right) \sim \frac{l(x)}{x^{\kappa}}, \; \kappa > 2, \tag{6.12}$$

where $l(x)$ is a function slowly varying at infinity. Thus, the random variable X does not satisfy the Cramer condition for $t > 0$.

We also assume that $A(x)$ meets the following (technical)

Assumption A1 for any fixed $h > 0$,

$$A(x+h) - A(x) = -\frac{h\kappa A(x)}{x}(1 + o(1)) \text{ as } x \to \infty. \tag{6.13}$$

Let $\eta = \eta(\pi)$ be the same as in (6.11).

Assumption A2 (i) there exists $\delta > 0$ such that, as $x \to \infty$,

$$\mathbb{P}\left(\eta > x\right) = o\left(\frac{1}{\log x \times \left(\log\log x\right)^{1+\delta}}\right);$$

(ii) as $n \to \infty$,

$$\mathcal{L}\left(f\left(1 - e^{-an}\right)|X > an\right) \xrightarrow{d} \mathcal{L}(\nu), \tag{6.14}$$

where ν is a random variable which is less than 1 with a positive probability.

Let $\mathcal{L}_{\pi}\left(Z_1 e^{-an}\right)$ be the law of $Z_1 e^{-an}$ under fixed π corresponding to $f(\pi; s)$. If $\mathbb{E}X < 0$ then (6.14) is equivalent to the following assumption, concerning convergence of random measures:

$$\mathcal{L}_{\pi}\left(Z_1 e^{-x} | X > x\right) \Longrightarrow \nu\delta_0 + (1-\nu)\,\delta_{\infty} \quad \text{as } n \to \infty, \tag{6.15}$$

where δ_0 and δ_{∞} are measures assigning unit masses to the corresponding points.

In what follows we assume that *the distribution of X is nonlattice*. The case when the distribution of X is lattice needs natural changes in the respective statements.

Define

$$f_{k,n}(s) = f_k(f_{k+1}(\ldots(f_{n-1}(s))\ldots)),\ 0 \le k \le n-1, \quad \text{and} \quad f_{n,n}(s) = s. \tag{6.16}$$

Theorem 6.5 *Assume that $a = -\mathbb{E}[X] > 0$ and conditions (6.12), (6.13) and Assumptions A1 and A2 are valid. Then the survival probability of the process $\{Z_n\}$ has, as $n \to \infty$, the asymptotic representation*

$$\mathbb{P}(Z_n > 0) \sim K\mathbb{P}(X > na), \tag{6.17}$$

where

$$K = \sum_{j=0}^{\infty} \mathbb{E}\left[1 - f_{0,j}(\nu)\right] \in (0, \infty), \tag{6.18}$$

and ν is a random variable that has the same distribution as the ν in Assumption A2(ii) and is independent of the underlying environment $\Pi = \{\pi_n\}$ (and consequently of $\{f_{0,j}\}$ for each $j = 0, 1, \ldots$).

It can be shown that if π_1 is either almost surely a Poisson distribution or almost surely a geometric distribution, then $\nu \equiv 0$ and the constant $K = \sum_{j=0}^{\infty} \mathbf{P}(Z_j > 0)$. In this case Theorem 6.5 has the following intuitive explanation. Let

$$U_n = \inf\left\{j : X_j > na\right\} \tag{6.19}$$

be the first time when the increment of the ARW exceeds na (the ARW has a big jump). Then the event $\{Z_n > 0\}$ is asymptotically equivalent to

$$\{U_n < n, Z_{U_n - 1} > 0\} = \cup_{j<n}\{Z_{j-1} > 0, U_n = j\}.$$

Now for each fixed $j \ge 1$,

$$\mathbb{P}(Z_{j-1} > 0, U_n = j) \sim \mathbb{P}(Z_{j-1} > 0)\mathbb{P}(X_1 > na),$$

and hence, not rigorously,

$$\begin{aligned}\mathbb{P}(Z_n > 0) &\sim \mathbb{P}(U_n < n, Z_{U_n-1} > 0)\\ &\sim \sum_{j=1}^{\infty} \mathbb{P}(Z_{j-1} > 0) \cdot \mathbb{P}(X_1 > na) = K\mathbb{P}(X_1 > na).\end{aligned}$$

In fact, one can show that the process $\{Z_k, k = 1, 2, \ldots\}$ survives up to a distant moment n owing to one big jump of the ARW happened at the very beginning of the evolution of the process; moreover, the big jump is accompanied by a population explosion which leads to survival.

We now give an example taken from [23] of a BPRE where ν is either positive and less than 1 with probability 1, or random with support not concentrated at 1. To this aim let ν be a random variable with values in $[0, 1-\delta] \subset [0, 1]$ for some $\delta \in (0, 1]$, and p and $q, p + q = 1, pq > 0$, be random variables independent of ν such that the random variable

$$X = \log(1-\nu) + \log\frac{p}{q}$$

meets the conditions $\mathbb{E}[X] < 0$ and (6.12). Define

$$f(s) = \nu + (1-\nu)\,\frac{q}{1-ps}.$$

Then $f'(1) = (1-\nu)\,p/q = e^X$, and

$$\begin{aligned}\nu \le f(1-e^{-an}) &= \nu + (1-\nu)\,\frac{q}{q+pe^{-na}}\\ &= \nu + (1-\nu)\,\frac{1}{1+(1-\nu)^{-1}\,e^{X-na}}\\ &\le \nu + e^{-(X-na)}.\end{aligned}$$

Since $\mathbb{E}[X] < 0$, we have for any $\varepsilon \in (0, 1)$,

$$\lim_{n\to\infty} \mathbb{P}\left(|f(1-e^{-an}) - \nu| \ge \varepsilon \,|\, X > na\right) \le \lim_{n\to\infty} \mathbb{P}\left(X - na \le -\log\varepsilon \,|\, X > na\right) = 0.$$

Therefore, as $n \to \infty$,

$$\mathcal{L}(f(1-e^{-an})|X > na) \xrightarrow{d} \mathcal{L}(\nu).$$

Let $\sigma^2 = Var(X)$. The following conditional functional limit theorem was proved in [23]:

Theorem 6.6 *Suppose that* $a = -\mathbb{E}[X] > 0$ *and* (6.12), (6.13) *and Assumptions A1 and A2 are valid. Then for any* $j \geq 1$,

$$\lim_{n\to\infty} \mathbb{P}(U_n = j | Z_n > 0) = \mathbb{E}(1 - f_{0,j}(\nu))/K.$$

Moreover,

$$\mathcal{L}\left(\frac{Z_{[nt]\vee U_n}}{Z_{U_n}\exp(S_{[nt]\vee U_n} - S_{U_n})}, 0 \leq t \leq 1 \middle| Z_n > 0 \right) \Longrightarrow \mathcal{L}\left(1, 0 \leq t \leq 1\right), \tag{6.20}$$

$$\mathcal{L}\left(\frac{1}{\sigma\sqrt{n}} \left(\log\left(Z_{[nt]\vee U_n}/Z_{U_n}\right) + nta\right), 0 \leq t \leq 1 \middle| Z_n > 0 \right) \Longrightarrow \mathcal{L}(B_t, 0 \leq t \leq 1), \tag{6.21}$$

and for any $\varepsilon > 0$,

$$\mathcal{L}\left(\frac{1}{\sigma\sqrt{n}} \left(\log\left(Z_{[nt]}/Z_{[n\varepsilon]}\right) + n(t-\varepsilon)a\right), \varepsilon \leq t \leq 1 \middle| Z_n > 0 \right) \\ \Longrightarrow \mathcal{L}(B_t - B_\varepsilon, \varepsilon \leq t \leq 1),$$

where B_t *is a standard Brownian motion.*

Therefore, after the population explosion at time U_n, the population drops exponentially at rate a, with a fluctuation of order $\exp(O(\sqrt{k}))$ with k the number of generations elapsed after the explosion. Moreover, it follows from (6.20) and the invariance principle that

$$\mathcal{L}\left(\log\left(Z_{[nt]\vee U_n}/Z_{U_n}\right) - (S_{[nt]\vee U_n} - S_{U_n}), 0 \leq t \leq 1 \middle| Z_n > 0\right) \Longrightarrow \mathcal{L}(0, 0 \leq t \leq 1),$$

and, therefore, at the logarithmic level the fluctuations of the population are completely described by the fluctuations of the associated random walk after the big jump. Moreover, we see by (6.21) that given $Z_n > 0$ the size of the population is not necessarily bounded! This phenomenon has no analogues for other types of subcritical BPRE's.

6.1.3.3 The Case $\beta \in (0, 1)$

In the previous subsection the favorable trajectories leading to the survival of the process up to moment n have one big jump exceeding $na(1-\varepsilon) = -n(1-\varepsilon)\mathbb{E}[X]$. For the case $\beta \in (0, 1)$ the situation is more delicate. Namely, the trajectories providing survival up to moment n have a big jump with values in the interval

$\left[bn - M\sqrt{n}, bn - M\sqrt{n}\right]$ where $b > 0$ is some parameter (see below) and M is sufficiently large.

Let us list the conditions which provide the appearance of this phenomenon.

Hypothesis A The distribution of X has density

$$p_X(x) = \frac{l_0(x)}{x^{\kappa+1}} e^{-\beta x},$$

where $l_0(x)$ is a function slowly varying at infinity, $\kappa > 2$, $\beta \in (0, 1)$ and, in addition,

$$\varphi'(\beta) = \mathbb{E}\left[Xe^{\beta X}\right] < 0. \tag{6.22}$$

To formulate one more basic condition we once again use nowadays classical technique of studying subcritical branching processes in random environment used, for instance, in [6–9, 17]. Denote by $\mathcal{F}_n$ the $\sigma-$algebra generated by the tuple $(\pi_1, \pi_2, \ldots, \pi_n; Z_0, Z_1, \ldots, Z_n)$ and let $\mathbb{P}^{(n)}$ be the restriction of $\mathbb{P}$ to $\mathcal{F}_n$. Setting

$$\gamma = \varphi(\beta) = \mathbb{E}\left[e^{\beta X}\right],$$

we introduce another probability measure $\mathbf{P}$ by the following change of measure

$$d\mathbf{P}^{(n)} = \gamma^{-n} e^{\beta S_n} d\mathbb{P}^{(n)}, \; n = 1, 2, \ldots \tag{6.23}$$

or, what is the same, for any random variable Y_n measurable with respect to $\mathcal{F}_n$ we let

$$\mathbf{E}\left[Y_n\right] = \gamma^{-n} \mathbb{E}\left[Y_n e^{\beta S_n}\right]. \tag{6.24}$$

Note that by Jensen's inequality and (6.22),

$$-a = \mathbb{E}[X] < \frac{\mathbb{E}\left[Xe^{\beta X}\right]}{\mathbb{E}\left[e^{\beta X}\right]} = \varphi'(\beta)/\varphi(\beta) = \mathbf{E}[X] = -b < 0.$$

Thus, under the new measure the BPRE is still subcritical and the ARW $\{S_n, \; n \geq 0\}$ tends to $-\infty$ as $n \to \infty$ with a smaller rate.

Hypothesis B There exists a random function

$$g(\lambda), \lambda \in [0, \infty), 0 < g(\lambda) < 1 \text{ for all } \lambda > 0,$$

and

$$\lim_{\lambda \to \infty} g(\lambda) = 0$$

such that, for all $k = 0, 1, 2, \ldots$

$$\lim_{y\to\infty} \mathbf{E}\left[f^k\left(\pi; e^{-\lambda/y}\right)|f'(\pi;1) = y\right] = \mathbf{E}\left[g^k(\lambda)\right]. \tag{6.25}$$

The following result was established in [12]:

Theorem 6.7 *If*

$$\mathbb{E}\left[-\log\left(1-\pi\left(\{0\}\right)\right)\right] < \infty, \quad \mathbb{E}\left[e^{-X}\sum_{k\geq 1}\pi\left(\{k\}\right)k\log k\right] < \infty \tag{6.26}$$

and Hypotheses A and B are valid, then there exist positive constants C_0 *and* C_1 *such that, as* $n \to \infty$

$$\mathbb{P}\left(Z_n > 0\right) \sim C_0\beta\gamma^{n-1}\frac{l_0(n)}{(bn)^{\kappa+1}} \sim C_1\mathbb{P}\left(L_n \geq 0\right). \tag{6.27}$$

We stress that $\gamma = \varphi(\beta) \in (0, 1)$ in view of $\varphi(0) = 1$ and (6.22).

In fact, one can show that

$$\mathbb{P}\left(Z_n > 0\right) \approx const \times \mathbf{P}(X_1 \in [bn - M\sqrt{n}, bn + M\sqrt{n}])$$

for n large enough. Thus, (6.27) once again confirms that in the subcritical regime the survival event is, as a rule, associated with the event when the ARW is bounded from below.

Note that to study the asymptotic behavior of the survival probability for the case $\beta = 0$ implying $\mathbf{P} = \mathbb{P}$, the condition was imposed that

$$\mathcal{L}\left(f\left(\pi; e^{-\lambda/y}\right)|f'(\pi;1) > y\right) \longrightarrow \mathcal{L}\left(\nu\right),\ y \to \infty,$$

where ν is a random variable being *independent* of $\lambda > 0$ and less than 1 with a positive probability. In this case the random walk $\mathbf{S}$ generated by the environment that provides survival up to a distant moment n should have a single big jump *exceeding* $(1-\varepsilon)\,an$ for any $\varepsilon > 0$. In the case $\beta \in (0, 1)$ the random walk generated by the environment, viewing under the measure $\mathbf{P}$ and providing survival up to a distant moment n should have a single big jump *enveloped* by $bn - M\sqrt{n}$ and $bn + M\sqrt{n}$ for a large constant M. This requires the validity of the conditions that are based on local properties of the random variable $f'(\pi; 1)$ and includes *dependence* of the limiting function in (6.25) on $\lambda > 0$.

Theorem 6.8 *Under the conditions of Theorem 6.7,*

$$\lim_{n\to\infty} \mathbb{E}\left[s^{Z_n}|Z_n > 0\right] = \Omega(s),$$

where $\Omega(s)$ is the probability generating function of a proper nondegenerate distribution supported on $\mathbb{Z}_+$.

We see that, contrary to the case $\beta = \rho_+ \wedge 1 = 0$ this Yaglom-type limit theorem has the same form as for the ordinary Galton-Watson subcritical processes.

For every pair $n \geq j \geq 1$, we define a tuple of random variables

$$W_{n,j} = \frac{1 - f_{n,j}(0)}{e^{S_n - S_j}} \tag{6.28}$$

and its limit

$$W_j = \lim_{n\to\infty} W_{n,j},$$

which exists by monotonicity of $W_{n,j}$ in n. We also define a random function $g_j : \mathbb{R}_+ \to [0, 1]$ such that

1. g_j is a probabilistic copy of the function g specified by (6.25);
2. $f_{0,j-1}$, g_j and $(W_{n,j}, W_j, f_k : k \geq j + 1)$ are independent for each $n \geq j$ (it is always possible, the initial probability space being extended if required).

We set

$$c_j = \int_{-\infty}^{\infty} \mathbb{E}\left[1 - f_{0,j-1}(g_j(e^v W_j))\right] e^{-\beta v} dv$$

and focus on the exceptional environment explaining the survival event by giving a more explicit result. For any $\delta \in (0, 1)$, let

$$\varkappa(\delta) = \inf\{j \geq 1 : X_j \geq \delta bn\}.$$

Theorem 6.9 *Under $\mathbb{P}$, conditionally on $Z_n > 0$, $\varkappa(\delta)$ converges in distribution to a proper random variable. Moreover, conditionally on $\{Z_n > 0, X_j \geq \delta bn\}$, the distribution law of $(X_{\varkappa(\delta)} - bn)/(\sqrt{n}\, VarX)$ converges to a law μ specified by*

$$\mu(B) = c_j^{-1} \mathbb{E}\left[1(G \in B) \int_{-\infty}^{\infty} \left(1 - f_{0,j-1}(g_j(e^v W_j))\right) e^{-\beta v} dv\right]$$

for any Borel set $B \subset \mathbb{R}$, where G is a centered gaussian random variable with variance $VarX$, which is independent of $(f_{0,j-1}, g_j)$.

The following two examples taken from [12] meet the conditions of Theorem 6.7.

Example 6.1 Let (with a slight abuse of notation) $\xi = \xi(\pi) \geq 0$ be an integer-valued random variable with probability generating function $f(\pi; s) = \mathcal{E}\left[s^{\xi(\pi)}\right]$. If $\mathbf{E}[\log f'(\pi; 1)] < 0$ and there exists a deterministic function $g(\lambda), \lambda \geq 0$, with

$g(\lambda) < 1, \lambda > 0$, and $g(0) = 1$, such that, for every $\varepsilon > 0$

$$\lim_{y\to\infty} \mathbf{P}\left(\pi : \sup_{0\le\lambda<\infty} \left|f\left(\pi; e^{-\lambda/y}\right) - g(\lambda)\right| > \varepsilon \,\middle|\, f'(\pi; 1) = y\right) = 0,$$

then Hypothesis B is satisfied for the respective subcritical branching process.

Example 6.2 If the support of the environment is concentrated on probability measures $\pi \in N$ such that, for any $\varepsilon > 0$

$$\lim_{y\to\infty} \mathbf{P}\left(\pi : \left|\frac{\xi(\pi)}{f'(\pi; 1)} - 1\right| > \varepsilon \,\middle|\, f'(\pi; 1) = y\right) = 0 \tag{6.29}$$

and the density $p_f(y)$ of the random variable $f'(\pi; 1)$ is positive for all sufficiently large y, then $g(\lambda) = e^{-\lambda}$. Condition (6.29) is satisfied if, for instance,

$$\lim_{y\to\infty} \mathbf{P}\left(\pi : \frac{Var\xi(\pi)}{(f'(\pi; 1))^2} > \varepsilon \,\middle|\, f'(\pi; 1) = y\right) = 0$$

for any $\varepsilon > 0$.

6.2 Subcritical BPRE Attaining a High Level

There are several papers in which the asymptotic behavior of the distribution of the random variables

$$\sup_{0\le n<\infty} Z_n \text{ and } Y = \sum_{k=0}^{\infty} Z_k$$

are investigated for subcritical BPRE's. The following result was established by Afanasyev in [3]. Assume that $\mathbb{E}[X] = \mathbb{E}[\log f'(\pi; 1)] < 0$ and there exists a positive number κ such that

$$\mathbb{E}\left[e^{\kappa X}\right] = 1 \tag{6.30}$$

and

$$\mathbb{E}\left[|X|\, e^{\kappa X} \log^+ X\right] < \infty,\ \mathbb{E}\left[e^{(\kappa-1)X} Z_1 \log(Z_1 + 1)\right] < \infty \tag{6.31}$$

and, in addition, if $\kappa \ge 1$ then there exists a $p > \kappa$ such that

$$\mathbb{E}\left[e^{(\kappa-p)X} Z_1^p\right] < \infty. \tag{6.32}$$

If the distribution of X is non-lattice then there exists a positive constant K_1 such that, as $x \to \infty$

$$\mathbb{P}\left(\sup_{0\le n<\infty} Z_n > x\right) \sim K_1 x^{-\kappa}.$$

Under the same conditions and for the case of fractional-linear probability generating functions $f_n(s)$ and $\kappa \in (0,2]$ it was shown in [19] that there exists a positive constant K_2 such that, as $x \to \infty$

$$\mathbb{P}(Y > x) \sim K_2 x^{-\kappa}.$$

An extension of the previous results to the multitype subcritical BPRE's counted by random characteristics was obtained in [22].

Afanasyev investigated in a number of papers properties of subcritical BPRE's assuming that $T_x = \min\{n : Z_n > x\}$ and x is large. A summary of his results presented in [4, 5] looks as follows.

Let $T = \min\{n : Z_n = 0\}$ and

$$d = \mathbb{E}\left[Xe^{\kappa X}\right], \qquad a = -\mathbb{E}[X].$$

If the conditions (6.30)–(6.32) are valid then, as $x \to +\infty$

$$\left(\frac{T_x}{\ln x}\,\middle|\, T_x < +\infty\right) \xrightarrow{\mathbb{P}} \frac{1}{d}$$

and

$$\left(\frac{T}{\ln x}\,\middle|\, T_x < +\infty\right) \xrightarrow{\mathbb{P}} \frac{1}{d} + \frac{1}{a}.$$

If, in addition,

$$\sigma^2 = \mathbb{E}\left[X^2 e^{\kappa X}\right] - d^2 < +\infty$$

then, as $x \to +\infty$

$$\mathcal{L}\left(\frac{T_{x^t} - t\ln x/d}{\sigma d^{-3/2}\sqrt{\ln x}},\ 0 \le t \le 1 \,|T_x < +\infty\right) \Longrightarrow \mathcal{L}(B_t, 0 \le t \le 1)$$

and, as $y \to +\infty$

$$\mathcal{L}\left(\frac{Z_{ty/d} - ty}{\sigma\sqrt{y/d}},\ 0 \le t < 1 \,|T_{e^y} < +\infty\right) \Longrightarrow \mathcal{L}(B_t, 0 \le t < 1).$$

Acknowledgements This work is supported by the RSF under a grant 14-50-00005.

References

1. Afanasyev, V.I.: Limit theorems for a moderately subcritical branching process in a random environment. Discrete Math. Appl. **8**, 55–72 (1998)
2. Afanasyev, V.I.: Limit theorems for an intermediately subcritical and a strongly subcritical branching process in a random environment. Discrete Math. Appl. **11**, 105–131 (2001)
3. Afanasyev, V.I.: On the maximum of a subcritical branching process in a random environment. Stoch. Process. Appl. **93**, 87–107 (2001)
4. Afanasyev, V.I.: High level subcritical branching processes in a random environment. Proc. Steklov Inst. Math. **282**, 4–14 (2013)
5. Afanasyev, V.I.: Functional limit theorems for high-level subcritical branching processes in random environment. Discrete Math. Appl. **24**(5), 257–272 (2014)
6. Afanasyev, V.I., Geiger, J., Kersting, G., Vatutin, V.A.: Criticality for branching processes in random environment. Ann. Probab. **33**, 645–673 (2005)
7. Afanasyev, V.I., Geiger, J., Kersting, G., Vatutin, V.A.: Functional limit theorems for strongly subcritical branching processes in random environment. Stoch. Process. Appl. **115**, 1658–1676 (2005)
8. Afanasyev, V.I., Boeinghoff, C., Kersting, G., Vatutin, V.A.: Limit theorems for weakly subcritical branching processes in random environment. J. Theor. Probab. **25**, 703–732 (2012)
9. Afanasyev, V.I., Boeinghoff, C., Kersting, G., Vatutin, V.A.: Conditional limit theorems for intermediately subcritical branching processes in random environment. Ann. Inst. Henri Poincare Probab. Stat. **50**, 602–627 (2014)
10. Athreya, K.B., Karlin, S.: On branching processes with random environments: I. Ann. Math. Stat. **42**, 1499–1520 (1971)
11. Athreya, K.B., Karlin, S.: On branching processes with random environments: II. Ann. Math. Stat. **42**, 1843–1858 (1971)
12. Bansaye, V., Vatutin, V.: On the survival probability for a class of subcritical branching processes in random environment. Bernoulli (2016, in print). http://www.bernoulli-society.org/index.php/publications/bernoulli-journal/bernoulli-journal-papers
13. Birkner, M., Geiger, J., Kersting, G.: Branching processes in random environment - a view on critical and subcritical cases. In: Deuschel, J.-D., Greven, A. (eds.) Interacting Stochastic Systems, pp. 269–291. Springer, Berlin (2005)
14. Böinghoff, Ch., Kersting, G.: Simulations and a conditional limit theorem for intermediately subcritical branching processes in random environment. Proc. Steklov Inst. Math. **282**, 45–61 (2013)
15. D'Souza, J.C., Hambly, B.M.: On the survival probability of a branching process in a random environment. Adv. Appl. Probab. **29**, 38–55 (1997)
16. Dyakonova, E., Geiger, J., Vatutin, V.: On the survival probability and a functional limit theorem for branching processes in random environment. Markov Proc. Relat. Fields **10**, 289–306 (2004)
17. Geiger, J., Kersting, G., Vatutin, V.A.: Limit theorems for subcritical branching processes in random environment. Ann. Inst. Henri Poincare Probab. Stat. **39**, 593–620 (2003)
18. Guivarc'h, Y., Liu, Q.: Proprietes asymptotique des processus de branchement en environment aletoire. C.R. Acad. Sci. Paris Ser. I Math. **332**, 339–344 (2001)
19. Kesten, H., Kozlov, M.V., Spitzer, F.: A limit law for random walk in a random environment. Compos. Math. **30**(2), 145–168 (1975)
20. Smith, W.L., Wilkinson, W.E.: On branching processes in random environments. Ann. Math. Stat. **40**, 814–827 (1969)
21. Vatutin, V.A.: A limit theorem for intermediate subcritical branching processes in a random environment. Theory Probab. Appl. **48**, 481–492 (2003)

22. Vatutin, V.A.: Polling systems and multitype branching processes in a random environment with a final product. Theory Probab. Appl. **55**, 631–660 (2011)
23. Vatutin, V.A., Zheng, X.: Subcritical branching processes in a random environment without the Cramer condition. Stoch. Process. Appl. **12**, 2594–2609 (2012)
24. Vatutin, V.A., Zubkov, A.M.: Branching processes II. J. Sov. Math. **67**, 3407–3485 (1993)
25. Vatutin, V.A., Dyakonova, E.E., Sagitov, S.: Evolution of branching processes in a random environment. Proc. Steklov Inst. Math. **282**, 220–242 (2013)

Part IV
Size/Density/Resource-Dependent Branching Models

Chapter 7
The Theorem of Envelopment and Directives of Control in Resource Dependent Branching Processes

F. Thomas Bruss

Mathematics Subject Classification (2000): 60J85, 60J05, 60G40

7.1 Introduction

For human populations certain features will probably never change. People want to live and be protected by the society in which they live. They will care for the future of their children. They will need food and resources and will have to think about production, distribution and consumption of resources. And then, they always seem to long for a higher standard of living.

Survival and increasing the standard of living may be the most natural and most important objectives of any human society. Therefore we single them out as two special hypotheses:

H1. Individuals of a human society would like to survive and see a future for their children.

H2. Individuals of a human society typically prefer a higher standard of living to a lower one.

The hypotheses **H1** and **H2** are frequently incompatible with each other. Therefore we establish the following rule of priority:

Priority: If **H1** and **H2** are in conflict with each other then **H1** is assumed to take priority before **H2**.

F.T. Bruss (✉)
Département de Mathèmatique, Université Libre de Bruxelles, CP 210, B-1050 Bruxelles, Belgium
e-mail: tbruss@ulb.ac.be

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_7

Is it possible to predict limiting society forms for possible developments of human societies just out of these two assumptions?

Before we answer any question about the effect of these hypotheses we must first clarify what exactly is meant by Hypothesis H1. Clearly, the survival probability should be strictly positive, and as large as possible. We will argue in the following sections that it is preferable not to quantify exactly with what probability the population wants to survive. As it is often the case for branching processes (see e.g. [11]), if the survival probability is strictly positive for an initial state of one ancestor, then under reasonable conditions it converges to one when the population grows. This will also be the case for most populations we will examine through our models.

Also, we have not yet defined what should be the rules to determine whether or not a given society respects the hypotheses H1 and H2. These rules will be established below in what we call the *society obligation principle.*

7.1.1 Society Obligation Principle

Since possible events like fertility changes, climate changes, unpredictable shortages of resources, war, or other catastrophes cannot be excluded, the objective to survive must be defined in terms of probability. Extinction is a tail event which cannot be determined in terms of finitely many generations, of course. We must therefore define first what we mean by saying that a society *respects* H1.

Definition 7.1 *Society obligation principle*:

A society is said to *respect* hypothesis H1, if it satisfies the following conditions:

(i) If the *currently* observed parameters *were to hold forever*, then the population would, with *strictly positive* probability, survive forever,

(ii) If (i) is not satisfied then the society must try to control immediately for survival, that is, to take all steps in its power to obtain, as quickly as possible, a *combination of parameters* which, if maintained forever, would yield under the same conditions a positive survival probability.

It is thus the set of the currently observed parameters which we define here to be the relevant measure for the change of rules (if necessary) since estimates for the future are risky. This safety precaution is seen as an integral part of the obligation.

As announced before, the preceding definition does not specify a minimum survival probability; it must just be strictly positive. Also, the definition contains no rules concerning hypothesis H2 on its own. Hence we maintain some freedom with respect to the average standard of living (which will be defined in Definition 7.4), and more generally, in what way parameters should be corrected, if necessary. This has the advantage that we do not have to suppose to know how much future societies would risk just to increase the standard of living.

7.1.2 Theorem of Envelopment and Related Results

Now let us return to the question whether it is possible to predict limiting society forms for possible developments of human societies just out of the hypotheses H1 and H2.

Surprisingly, the answer is yes if human societies are modelled by sequences of so-called *resource dependent branching processes*, in short RDBPs. In particular, populations which are guided by the implications of H1 and H2 will live in the long run in, what we call, an *envelope.*

This follows from the *Theorem of Envelopment of Societies in Resource Dependent Branching Processes* [7] and from earlier results in [5, 8] surrounding this theorem. Wajnberg [19] and others refer to these results, all taken together, as the Bruss-Duerinckx Theorem. The combination of these results will show how the survival criteria of the societies forming the mentioned *envelope* of possible societies can be explicitly computed.

7.2 Main Definitions and Results

To streamline our discussion, we first recall the construction of RDBPs as well as those results relating the envelopment theorem which will be needed.

A RDBP is a discrete time branching process modelling a population in which individuals have to work in order to be able to live and to reproduce. It involves production and consumption of resources, reproduction of individuals, decision rules (policies) for distributing resources and a means of interaction between the individuals and the society in which they live. The process counting the (remaining) individuals is denoted by (Γ_n). We make the following assumptions:

- **A1** All individuals which are *admitted* to reproduce within the population do so independently of each other according to the same law $\{p_k\}_{k=0}^{\infty}$. The variable $D_n(\Gamma_n)$ denotes the random number of descendants generated by Γ_n individuals for generation $n+1$. The mean asexual reproduction per individual $m = \sum_{k=1}^{\infty} kp_k$ is supposed to be finite.
- **A2** Individuals present in the population inherit resources, claim and consume resources from the preceding generation and create themselves resources for the next generation. The net outcome of all these activities is summarized as individual *production* of resources, and its finite mean is denoted by r.
- **A3** All production (resources, goods, services) is thought of as being evaluated in terms of money and to go into a common pool called resource space. The resource space available in the nth generation is denoted by $R_n := R_n(\Gamma_n)$.
- **A4** Each individual claims an attribution of resources for its own consumption, simply called *claim.* All claims are supposed to be i.i.d random variables with an absolute continuous distribution function F. Here X_k^j denotes the claim of the jth individual in generation k. The mean claim $\mu := \int_0^{\infty} xdF(x)$ is supposed to be finite.
- **A5** A policy π_n is a rule to distribute resources among individuals of generation n as long as the available resource space R_n suffices. Formally it is a function defined of the permutation group of the list of claims brought forward in generation n. The output of

π_n is a counting function Q^{π_n} with values in $\{0, 1, 2, \cdots, D_n(\Gamma_n)\}$ which depends also on $R_n(\Gamma_n)$ and the list of claims. Its value determines the number of current descendants that will remain in the population until the next reproduction time.

With these assumptions we can now give the formal definition of a RDBP:

Definition 7.2 Under the Assumptions A1–A5, let $(\Gamma_n)_{n=0}^{\infty}$ be a discrete time stochastic process defined by

$$\Gamma_0 = 1 \tag{7.1}$$

and recursively,

$$\Gamma_{n+1} = Q^{\pi_n}\left(D_n(\Gamma_n), \left(X_n^{k=1,2,\cdots,D_n(\Gamma_n)}\right), R_n(\Gamma_n)\right), \ n = 0, 1, \cdots \tag{7.2}$$

Then $(\Gamma_n)_{n=0}^{\infty}$ is called a Resource Dependent Branching Process.

Definition 7.3 If we start (Γ_n) with $k \in \{1, 2, \cdots\}$ individuals at time 0, that is with initial condition (7.1) being replaced by $\Gamma_0 := \Gamma_0(k) = k$, then the resulting process satisfying the corresponding recurrence relation (7.2) is denoted by $(\Gamma_n(k))$. Hence, for $n = 0, 1, \cdots$,

$$\Gamma_{n+1}(k) = Q^{\pi_n}\left(D_n(\Gamma_n(k)), \left(X_n^{j=1,2,\cdots,D_n(\Gamma_n(k))}\right), R_n(\Gamma_n(k))\right) \tag{7.3}$$

and $(\Gamma_n) \equiv (\Gamma_n(1))_{n=0}^{\infty}$.

Definition 7.4 The random variable

$$\frac{R_n(\Gamma_n(k))}{\Gamma_{n+1}(k)} \tag{7.4}$$

is called the random average standard of living of the process $(\Gamma_n(k))_{n=0}^{\infty}$ in generation $n + 1$.

Interpretation: Individuals inherit and consume resources; they also create resources, which they may consume and/or save for the next generation. They have individual resource claims and an option of protest or non-cooperation, if the current policy does not satisfy their claims. In the RDBP model of Bruss and Duerinckx [7], an individual whose claim is not completely satisfied, leaves the population before reproduction. Hence in their model *emigration* is seen in its strict sense, and meant to be the interacting tool for individuals to express discontent with the society.

Emigration can however be understood in a broader sense as a means of protest against society. For instance, an individual can also be understood to *emigrate* if it decides to make no net-contribution in production to the society and to have deliberately no children.

The random claims of individuals are thought of as being the outcome of individual desires and ambitions as well as individual strength to defend these desiderata against competition from the other descendants of the same generation. We note that the actual average consumption is not the same as an average claim, and that Definition 7.4 describes in (7.4) an average theoretical *richness* rather than what consumers may see as an average standard of living. With (7.3) this is meaningful also for smaller n if we suppose k to be large. If the limiting standard of living exits as $n \to \infty$, it is typically the same for all initial values k.

7.3 Special Societies

In the following we single out two specific societies, and this for two reasons. Firstly, they belong to a class of RDBPs for which, using a result of [8] on stopping times of order statistics, we can determine the survival criteria. Secondly, these policies play a particular role from an ideologic point of view, polarising weakest and strongest individuals.

The first one, the so-called *weakest-first society* (*wf*-society) always serves the smallest claims first, as long as resources suffice. This means that the *wf*-society is the most conservative spender of resources and therefore, with all other parameters kept constant, will allow for more individuals to stay than any other society. The extinction probability should therefore be the smallest possible, and if it survives, it should grow more quickly that any other society.

This intuition is correct as the following results show. For the notation recall the Assumptions A1–A5.

Theorem 7.1 *The wf-society has, for any initial state* $k > 0$, *a positive probability of survival, if either*

$$r > m\mu \tag{7.5}$$

or, if (7.5) is not satisfied, if

$$mF(\tau) > 1, \text{ where } \int_0^\tau x dF(x) = \frac{r}{m}. \tag{7.6}$$

Remark 7.1 The condition $r > m\mu$ in (7.5) says that the total expected production of a randomly chosen individual, i.e. heritage plus creation minus consumption, is larger than the expected sum of claims of its descendants. This implies, in the long run, the existence of more than enough resources for everyone. This in contrast to the typical economic principle of shortness of resources. Therefore we confine for the remainder of our paper our interest to the case $r \leq m\mu$.

For the same fixed parameters and initial values, let $(W_n)_{n=0}^{\infty}$ be the *wf*-process, and $(\Gamma_n)_{n=0}^{\infty}$ an arbitrary RDBP. Then we have (see [7]):

Theorem 7.2

$$\text{For all } n \in \{1, 2, \cdots\} : \; P(\Gamma_n \leq W_n) = 1. \tag{7.7}$$

The *wf*-society may moreover be seen as a *safe-haven* society with respect to the risk of distinction, because it has also the following property:

Theorem 7.3 *Let* $(W_n)_{n=0}^{\infty}$ *be a wf-process and let* q_{wf} *be its extinction probability under the initial condition* $W_0 = 1$. *Then for all* $k \in \{1, 2, \cdots\}$

$$\lim_{n\to\infty} P\left(W_n = 0 \middle| W_0 = k\right) = P\left(\lim_{n\to\infty} W_n = 0 | W_0 = k\right) \leq q_{wf}^k. \tag{7.8}$$

Remark 7.2 The state 0 is absorbing for all (Γ_n), since there is no immigration in RDBPs. This implies the equality $\lim P(\cdot) = P(\lim(\cdot))$ here as well as in some other places in this paper. Also, we note that (7.8) implies, unless $q_{wf} = 1$, that the probability of extinction goes very quickly to 0 as k increases, and, together with (7.7), that there is no better way to escape extinction. This is why we call it the safe-haven property.

7.3.1 Limits of Growth Parameters

An arbitrary RDBP need not have an asymptotic growth rate. As an example we may think of an obscure society which chooses very different rules to distribute resources according to the number of descendants in generation n being odd or even. However, if it does have an asymptotic growth rate then the following relationship will hold:

Theorem 7.4 *Let* (Γ_n) *be an arbitrary RDBP, and let* (W_n) *be the wf-process. Further let* τ *be defined by (7.6). Then*

$$P\left(\lim_{n\to\infty} \frac{W_{n+1}}{W_n} = mF(\tau) \,\middle|\, W_n \to \infty\right) = 1. \tag{7.9}$$

Moreover, for any arbitrary RDBP (Γ_n)

$$\text{If } P\left(\lim_{n\to\infty} \frac{\Gamma_{n+1}}{\Gamma_n} = \ell \,\middle|\, \Gamma_n \to \infty\right) = 1 \text{ then } \ell \leq mF(\tau). \tag{7.10}$$

Remark 7.3 Note that if the *wf*-policy allows for a positive survival probability $1 - q_{wf}$ then, according to (7.8), any society will most probably survive if it changes (whenever necessary) into the direction of a *wf*-society. As said before, according to (7.7) and (7.10) no other RDBP will do better with respect to survival.

7.3.2 Strongest-First Society

It is natural to expect that the *sf*-society which wastes resources more than any other society and can thus retain less individuals would then always lag behind any other society. Interestingly, this is wrong, and an explicit counterexample is given in [7], Sect. 6.2. In other words, one cannot hope for similarly strong comparison results for the *sf*-process as in (7.7) for the *wf*-process (W_n). Under the additional condition of bounded individual claims [7] proved however the following result:

Theorem 7.5 *Let θ be the unique solution (see Remark 7.1) of the equation*

$$\int_{\theta}^{\infty} x dF(x) = \frac{r}{m}, \tag{7.11}$$

and let $(S_n)_{n=0}^{\infty}$ denote the number of individuals in the sf-society in generation n. Then, for any distribution F with bounded support,

$$P\left(\lim_{n\to\infty} \frac{S_{n+1}}{S_n} = m(1 - F(\theta)) \,\middle|\, S_n \to \infty\right) = 1. \tag{7.12}$$

Interpretation: Recall that the parameter $mF(\tau)$ plays an important role in Theorem 7.1, namely no society can survive unless $mF(\tau) > 1$. The parameter $m(1-F(\theta))$ defined through (7.11) plays also a distinguished role. If this parameter exceeds 1 then the population, once it has reached a sufficient size, can afford to select any society form it wants without risking extinction.

7.3.3 The Envelope of Human Societies

We mentioned subtleties concerning the *sf*-process in the sense that, although (7.12) holds, it cannot serve as a uniform lower bound (taken over all possible societies) for effectives of a RDBP. The more surprising is therefore the fact that, as Bruss and Duerinckx [7] have shown, the essence of our intuition is correct after all. The *wf*-society has the largest survival probability of all societies, and as soon as populations have a sufficiently large size, the *sf*-society will have the smallest survival probability of all. Here is the precise formulation:

Theorem 7.6 *Let $(W_n(k))$ be the wf-process, $(S_n(k))$ be the sf-process, and $(\Gamma_n(k))$ be a process with an arbitrarily chosen policy, all with the same initial value k. Then*

$$P\left(\lim_{n\to\infty} S_n(k) \le \lim_{n\to\infty} \Gamma_n(k) \le \lim_{n\to\infty} W_n(k)\right) \to 1, \text{ as } k \to \infty. \tag{7.13}$$

Moreover, the right-hand-side inequality holds almost surely for all $k \in \{1, 2, \cdots\}$.

We conclude that the *wf*-society, respectively the *sf*-society will, in the long run, envelope (bound from above, respectively, from below) any arbitrary RDBP. This is why Theorem 7.6 earns the name *Theorem of envelopment*.

7.4 Macro-Economic Interest Versus Tools

Theorem 7.6 displays interesting macro-economic facts. Bruss and Duerinck [7] dare to call them fundamental because the theorem of envelopment spelled out in (7.13) is true, although the intuition, why it should be true, is not only misleading but in general false in one important direction. Having said this it is, as indicated in the Introduction, not always clear in how far interesting facts can be turned into interesting tools, that is, to produce concrete advice for societies following specified objectives.

Now, there are indeed instances where this is the case, that is, where the implications of the envelopment theorem yield immediate directives. Two relatively recent decisions (2013, respectively 2014) of the German grand-coalition government, widely discussed in the media, were chosen in [6] to illustrate this. In the latter, the application is almost straightforward, and the conclusion is that both decisions of the German government should be reconsidered.

There is nothing special about the choice of Germany, except that two facts coincide: firstly, Germany is a country with a particularly low natality rate, and secondly, the mentioned legislations concern earlier retirement and a change in the minimum wage policy. It is intuitive that reducing the age of retirement may cause financial problems for a state. However, it may come as a surprise even to specialists that such a step may speed up extinction, and this is true under very general conditions.

We shall add more to these examples in Sect. 7.7.

Now, if we speak in the above examples of special coincidences which make conclusions rather straightforward, it is natural to ask to what extent control advice is at reach for more general circumstances. For instance, is it possible under certain constraints to find a sufficiently large class of RDBPs (embeddable in our global model) which allow for *sub-envelopes* of interest which can be *computed* and will then lead to new directives of control? This question, for which Bruss and Duerinckx have already partial answers (work in progress) should be the focus of interest for whoever is interested in modelling human societies.

7.5 Local Models and Control

Since in reality many variables within a society depend on each other, a realistic model for the evolution of a population process cannot be obtained through a single RDBP unless one would allow for many more dependencies. This however would render RDBPs intractable. Moreover, this would not be skilful either because many future developments and changes in parameters are unpredictable. Hence it would not be convincing to fix, once and for good, a certain RDBP and to expect this process to model the evolution of a population in a realistic way. The idea emerging from these arguments is to consider evolution as a random sequence of locally updated RDBPs.

But what should we suggest as the *law* of this sequence, that is, what should determine the choice of the local model in each term of the sequence?

It is here that the society obligation principle assumes its role: We imagine that the society looks at the current rates of natality, current rates of productivity, at the way how claims of resources have developed, etc. Now, several parameters can be influenced by the rules applied by society and updating will intervene. For instance, society may encourage individuals to have more children, or consume less on the average, or curtail larger claims, or, if possible, impose a higher productivity.

The guidelines of control are determined by what the society wants, and here the basic hypothesis H1 (survival) stands out by definition. Increasing the standard of living (H2) is desirable, but, exactly as the society obligation principle requires, H1 has first priority.

7.6 The Global Model

Our global model for the development of human society is therefore a sequence of RDBPs which is in each generation (locally) compatible with the society obligation principle. At each control time the society takes its decisions based on what would happen if they went on as before. Here it is natural to think of generation time points $1, 2, \cdots$ as being the typical times of control, but the choice of other times (shorter periods, or times of census) are compatible with our model. Note that each local model is now for a shorter period only. Assumptions A1–A5 are therefore easier to justify.

In the global model we do not want to add further constraints to the society obligation principle. For instance, if the society is satisfied with the currently observed parameters, the update may not lead to a new model. Therefore we cannot predict how a population will develop. However, it suffices to understand the local RDBPs in order to find non-trivial bounds within which populations can develop.

Before we advertise the global model any further we look first in more detail into the simplifying assumptions we made for the local models, i.e. into the assumptions we made for the RDBPs.

7.6.1 Bisexual and Asexual Reproduction

In the definition of a RDBP no reference is made to the reproduction mechanism, although it seems necessary that the model be consistent with bisexual human reproduction. We explain here why we can confine our interest to asexual reproduction.

We see from the definition that RDBPs are Markov processes, although complicated ones. Since 0 is, by definition, an absorbing state for all RDBPs, they must either die out or explode. Hypothesis H1 has priority before an increase of the standard of living, and thus the interesting RDBPs are only those which have a positive survival probability.

The definition of a RDBP does not involve bisexual reproduction and hence no mating functions. We simply speak of individuals and of their independent reproduction defined by the law $\{p_k\}_{k=0}^{\infty}$ (see A1). For human populations it is reasonable to assume the reproduction mean m to be finite, and, as usual, that the conditions $0 < p_0 < 1$ and $p_0 + p_1 < 1$ are always satisfied. It is understood that this Galton-Watson process type reproduction is only assumed for the *reproduction times*. Clearly, these processes themselves are submitted to heavy internal dependencies and are in general hardly comparable with Galton-Watson processes.

This simplification is justified by being allowed to confine interest to the limiting behaviour of reproduction and also by the existence of the so-called *average reproduction rate per mating unit* introduced in [5]. For a discrete time bisexual population process counting *mating units,* $(Z_n)_{n=0}^{\infty}$, say, it is defined by

$$\tilde{m} = \lim_{k\to\infty} \frac{1}{k} E(Z_{n+1}|Z_n = k). \tag{7.14}$$

This definition gave a useful answer to an interesting question put forward in [12]. See also [10]. Nowadays this limiting rate is usually called *average reproduction rate* in the literature about bisexual population processes.

In [7], the existence of the limiting average reproduction rate is taken for granted for human populations living under the same society form and under the same conditions. Indeed, this can be justified for human populations for several obvious reasons. Moreover, it was nice to see in the work presented at the workshop by Prof. Mota (the slides of this talk, and all the following quoted in this paper can be downloaded from http://branching.unex.es/wbpa15/index.htm, clicking on Abstract menu) that assuming the existence of (7.14) finds support in many different forms, and also for populations other than human populations.

7.6.2 RDBPs and φ-Branching Processes

The presentation of Prof. G. Yanev was a good reminder that there is some connection between RDBPs and the so-called φ-branching processes already introduced by

Sevastyanov and Zukov in the early seventies. Indeed, the function Q^{π_n} defined in a RDBP determines the number of individuals in the nth generation which are allowed to reproduce. Interesting results by Yanev [21], as well as a generalisation of these without independence assumptions by Bruss [4] may therefore catch interest in the domain of RDBPs. However, it seems difficult to exploit this connection easily when we pass from the local model of a RDBP to the global model defined as a sequence of RDBPs. As far as the author feels, only a sequence of updated models has a chance to model approximately the evolution of a human society.

In that respect, population size dependent branching processes are closer to RDBPs, as we shall see below.

7.6.3 RDBPs and Population-Size Dependent Branching Processes

The presentations of Prof. Jagers and of Prof. Klebaner showed that, for certain questions, RDBPs may directly profit from their results on population-size dependent reproduction. This concerns in particular results about what Jagers and Klebaner and others call the *quasi-stationary phase.*:

Let (Z_n) be a (asexual) branching process with initial state Z_0 and reproduction rates which are supercritical below a certain threshold K, and subcritical above K. See [16]. Let the reproduction rate be $m(k)$ if the current size is k and suppose moreover that for some neighbourhood of K, $V(K)$, say,

$$m(k)_{k\uparrow K} \downarrow 1 \text{ and } m(k)_{k\downarrow K} \uparrow 1, \tag{7.15}$$

for all $k \in V(K)$. Let $Z_0/K = 1$ and, for fixed $0 < \epsilon < 1$, let

$$\sigma_\epsilon^{(K)} = \inf\left\{t \geq 0 : \left|\frac{Z_t}{K} - 1\right| > \epsilon\right\} \tag{7.16}$$

Then there exists positive constants $c_1(\epsilon)$ and $c_2(\epsilon)$ such that

$$E\left(\sigma_\epsilon^{(K)}\right) > c_1(\epsilon)e^{c_2(\epsilon)K}. \tag{7.17}$$

Indeed, we see that the hypotheses H1 and H2 and the society obligation principle play perfectly in the hands of the results of Jagers and Klebaner: If the effective reproduction rate is greater than one, then the society is not afraid of the risk of extinction and will, according to H2, increase its standard of living. The easiest way to do this is to serve more bigger claims. This policy will reduce the number of those individuals which will be able to stay, i.e., reduce the effective reproduction rate. However, if the latter falls below one, then hypothesis H1 takes over. The society obligation principle will force the society to increase its effective reproduction rate.

We have argued already by the average reproduction rate defined in (7.14) that bisexual reproduction makes no difference to asexual reproduction as far as asymptotic phenomena are concerned. Now, in the population-size dependent setting it is the *law of reproduction* which may fluctuate, whereas in a sequence of RDBPs, it is the *effective reproduction rate* which may fluctuate through the control of resource production and/or claims and/or natality rates. However, the reasons *why* certain laws of reproduction are chosen in each generation should be irrelevant because, like in the Jagers-Klebaner model, those individuals which *do reproduce* in RDBPs, do so independently of each other.

Therefore, if in the global model the control is *directly effective* each time already in the next generation, a result in the spirit of Theorem 7.1 must hold. Criticality plays here a central role in the global model, and thus in RDBPs, because the interplay of H1 and H2 makes near-criticality an attractive *policy* for societies. We should mention that the results in [1] and in [2] should also merit interest for RDBPs in that respect.

The author wonders about a question dressed in terms of a population size dependent process and related to the above cited work of Jagers and Klebaner: Let δ be a small positive real number, and let $V(\delta) = [1 - \delta, 1 + \delta]$. V is thought of as being a target interval for the (effective) reproduction mean of a population size dependent process.

> Question: Suppose that the current state is K and that the current (effective) reproduction mean equals ρ. Suppose further that the random number of generations necessary to bring the reproduction mean back into $V(\delta)$ has (waiting time) distribution $W_{\delta,\rho}(w)$. Under which general conditions on $W_{\delta,\rho}(w)$ does
>
> $$E\left(\sigma_{V(\delta)}\right) > \tilde{c}_1(\delta)e^{\tilde{c}_2(\delta)K}, \tag{7.18}$$
>
> hold for corresponding constants $\tilde{c}_1(\delta)$ and $\tilde{c}_2(\delta)$, where now $\sigma_{V(\delta)}$ denotes the total sojourn time of the reproduction mean in V.

Remark 7.4 The author sees the interest for this question as being obvious for RDBPs. The assumption (7.15) can be assumed to hold and may simplify the proof. It is also realistic to suppose that the waiting time mean is uniformly bounded away from 0.

7.7 Understanding Unexpected Implications

Bruss and Duerinckx [7] comment in Sects. 8.1–8.3 on classical society forms, including Mercantilism, Marxism, Leninism and Capitalism. The following comments are more specific and more applied in the sense that they allow, at least in certain cases, to derive definite advice how control parameters of a society having specific objectives. Bruss [6] examines two examples of recent legislation in Germany. We take these examples as guidelines of our approach.

It is obvious that any society considering a change of legislation concerning the standard age of retirement must keep financial constraints in view. Values are created by production, and if the total expected time of production of individuals change, the expected output will change, although not necessarily in the same proportion.

In the language of RDBPs the expected output is the expected contribution r to the currently available common resource space R, say. We recall that we think of all types of production as well as the total resource space as being measurable in monetary units. If r will go up or down, R will do so accordingly. If a society considers legislation allowing for an earlier retirement then it must consider the consequences. If a smaller R should suffice to pay the same pension to the same number of people working less than before then, as it seems, there should something wrong.

Now, a priori, nothing need be wrong. There is the possibility that R is more than sufficient, that is, that the previously paid pensions were not as high as they could have been and leave a large reserve of funds. Viewing hypothesis H2 we must conclude however, that this should rather be a short-time phenomenon because typically people will vote for those politicians advertising and promising a higher standard of living and who will thus spend the money. Finally these voters will succeed in getting what is possible. But then, when there are no reserve funds, there must be something wrong indeed. Decreasing the expected output r should lead (with life expectancies going up rather than going down) to smaller pension payments. Thus reducing the age of retirement is clearly a question of how to finance it.

7.7.1 Hidden Phenomena

It may come as a surprise that, apart from these financial problems, reducing the age of retirement may accelerate a decrease of the effectives of a population and thus increase the probability of extinction. For RDBPs at least, this is, as the next theorem will show, the case if the natality rate is subcritical, that is if $m < 1$.

Theorem 7.7

(i) *If the natality rate m of a RDBP (Γ_n) is smaller than 1 and the average production r does not exceed $m\mu$ (where μ is the average claim) then, for m fixed, a reduction of the rate r implies an increase of the extinction probability of (Γ_n).*

(ii) *For a fixed average production r the extinction probability of an arbitrary RDBP (Γ_n) decreases as m increases.*

Proof

(i) Remember that, with respect to H1, the society obligation is intrinsic in the definition. Let now $0 < m/r \leq \mu$ be fixed. Recall that F, the cumulative distribution function of the random claims, was supposed to be absolutely continuous, and it is monotone increasing by definition. Hence there exists a unique solution $\tau := \tau[r,m]$ of the equation

$$\int_0^{\tau[r,m]} x dF(x) = \frac{r}{m}. \tag{7.19}$$

According to Theorem 7.1, (7.6), (Γ_n) is bound to get extinct since $mF(\tau) < 1$.

Now let $G(r,m) = mF(\tau[r,m])$. Using again the continuity of F we see from (7.19) that $\tau[r,m]$ is by its definition differentiable in r. It follows that

$$\frac{\partial G(r,m)}{\partial r} = m\frac{dF(\tau[r,m])}{d\tau[r,m]}\frac{\partial \tau[r,m]}{\partial r} \geq 0, \tag{7.20}$$

because $m\partial F(\tau)/\partial \tau$ on the right-hand side of (7.20) is clearly nonnegative, and the factor $\partial \tau/\partial r$ as well according to (7.19). Hence reducing r implies reducing the growth rate $mF(\tau[r,m])$. This is the growth rate of the *wf*-society. The priority rule of H1 before H2 together with the society obligation forces the society governing the process (Γ_n) however to try to realize this growth rate $mF(\tau[r,m])$ which, according to (7.10) in Theorem 7.4, is best possible. Hence statement (i) of Theorem 7.7 is proved.

(ii) Although this statement seems evident, the proof takes slightly more than the proof of (i). We look again at the definition of $\tau[r,m]$ in the integral equation (7.19). Since the relevant parameter m is in the denominator on the right-hand side, $\tau[r,m]$ is now a decreasing function of m. As $\tau[r,m]$ is differentiable in m for all $m > 0$ we have $\partial\tau(r,m)/\partial m \leq 0$. Clearly

$$\int_0^{\tau} x dF(x) \leq \int_0^{\tau} \tau dF(x) = \tau F(\tau). \tag{7.21}$$

Hence from (7.19)

$$F(\tau[r,m]) \geq \frac{r}{m\tau[r,m]}. \tag{7.22}$$

The partial derivative of both sides of (7.19) with respect to m yields, on the one hand,

$$\tau[r,m] f(\tau[r,m])\frac{\partial \tau[r,m]}{\partial m} = \frac{-r}{m^2}, \tag{7.23}$$

where f denotes the corresponding density of random claims. On the other hand we obtain

$$\frac{\partial G(r,m)}{\partial m} = F(\tau[r,m]) + mf(\tau[r,m])\frac{\partial \tau[r,m]}{\partial m}. \tag{7.24}$$

This implies, using (7.22) and (7.23) that the right-hand side of (7.24) is positive so that $\partial mF(\tau[r,m])/\partial m$ is also positive.

This proves (ii) and completes the proof of Theorem 7.7. □

Remark 7.5 One should note that Theorem 7 could be stated without knowing more about F and the parameters r and m. All we needed was that F is an absolute continuous distribution function, and for (i), that $r \leq m\mu$.

If we let vary both parameters m and r at the same time then we need to know more, of course, to make conclusions. Indeed, now we would have to know the interplay between r and m, in other words, to know where the total differential

$$\Psi_{r,m}(dr, dm) := \frac{\partial mF(\tau[r,m])}{\partial r} dr + \frac{\partial mF(\tau[r,m])}{\partial m} dm \tag{7.25}$$

will vanish if r varies with m and vice versa.

Example 7.1 (Earlier Retirement) In 2014 the German government passed a law reducing the age of official retirement from 63 to 65 years. Some constraints do apply so that the average working time of people is unlikely to decrease by the two full years. However, it will definitely decrease, and thus with all other parameters staying equal, the average production rate r will decrease. Since the current German natality rate is only about $m \sim 0.71 < 1$ per individual (the lowest in Europe and among the lowest in the world), case (i) of Theorem 7.7 applies. We conclude that this legislation not only poses the problem of finding funds to finance the earlier retirement but, more seriously, even accelerate the de-population of Germany.

Of course, it is the extremely low natality in Germany which is the main problem, but natality is what it is for the moment, and no quick increase of the birth rate can be expected. Therefore r is so important. For foreign media which usually do not speak of the German natality problem but rather prefer to hint to productivity as a good example for others, it may sound like a paradox that German productivity may be a problem. But, as we have just seen, it is not a paradox. Reducing the age of retirement is in contradiction to H1 and to the society obligation principle. It should be urgently re-considered.

Remark 7.6 Example 7.1 would of course apply similarly to any country with natality rate $m < 1$. As far as the author could find out, there was (fortunately) no other country in recent decades with $m < 1$ where the government would propose a reduction of the retirement age. On the contrary, countries with a better natality, as for instance Belgium, envisage increasing the legal age of retirement. Ironically, some Belgian media hint to Germany to find public support for their project, saying that Germany plans to lift the retirement age up to 67. This is confusing media information. Indeed, certain groups of people in Germany will be able to work until the age of 67 or possibly longer. As we all know, a truth being valid on a subset may be very misleading on the whole set which counts.

7.7.2 Discussion

One could argue that the conclusions in Example 7.1 are based only on asymptotic results and may not be valid. Our answer would be that with such a low individual reproduction rate of 0.71, the asymptotic truth will quickly coincide with the truth.

Furthermore one could argue that the parameters m and r could change into a positive direction so that the problem of an accelerated extinction would disappear. But then we must object that we cannot know this and thus society should follow the rule as determined by the society obligation principle. The latter requires that decisions must be based on the currently observed parameters.

Finally, one might object that it is not clear that the society obligation principle must be imposed. Here our defense is bound to be somewhat weaker. We cannot *prove* that the society obligation principle is compulsory in an adequate society model because it has an axiomatic component. Indeed, if we have to decide what it means that a society respects hypothesis H1 and gives H1 priority before H2, then we must create a contradiction-free framework which defines what exactly this means. This is what we have done in our *principle* where the currently observed parameters are defined as the relevant ones. From the point of reliability this is arguably the best framework. However, clearly, a contradiction-free framework need not be unique.

Example 7.2 (Minimum Wage) Minimum wage legislation has been on the agenda of several governments in recent years. Does the introduction of a nationwide minimum wage increase productivity, or does it rather reduce it? Further, if it is true that a minimum wage policy will reduce the number of available jobs in the working class, would it then also be true that overall natality would go up because jobless people may have more time for other things?

There are plenty of discussions about this everywhere. The message RDBPs can tell is that, once one believes to have understood the interaction of m and r then any society which respects H1 and H2 must study the behaviour of the equation $\Psi_{r,m}(dr, dm) = 0$. With $f(x) = dF(x)/dx$ the left-hand side of this equation becomes

$$\left(f(\tau[r,m])\frac{\partial \tau[r,m]}{\partial r}\right) m\, dr + \left(F(\tau[r,m]) + mf(\tau[r,m])\frac{\partial \tau[r,m]}{\partial m}\right) dm \quad (7.26)$$

so that we would have to know the exact form of the density of claims f.

Understanding the *interplay* of the two parameters m (natality) and r (production), as well as f (linked with consumption) is important for any society respecting H1 before H2 because the effective reproduction mean is essential. Moreover, if the safe-haven rate $mF(\tau[r,m])$ approaches 1, then much may depend on the way it approaches 1, as exemplified by results in [17] for size-dependent BPs.

7.7.3 Immigration

In the definition of a RDBP emigration plays a fundamental role in the model as a means of interaction between individuals and society. Immigration is not incorporated however, at least not so far. This contrasts the numerous interesting contributions concerning and/or including immigration for other BPs as we have heard in the expositions in the workshop by Prof. Ispàny Prof. Pap (see also [13, 14]) as well as Prof. Mota [18]. If one thinks of the increasing number of questions immigration raises for *human society*, this is a weakness of RDBPs.

How could one overcome this weakness? One may be tempted to try to mimic immigration by a temporarily increased reproduction rate. See in that context the subsection *Sequence of branching processes with immigration* of the presentation of Prof. Ispàny as well as the earlier result in [20]. Indeed, recalling again the results of [15] for population size dependent BPs we know that population size dependence is quite a tractable setting.

However, there are at least two reasons why one has to be careful. Firstly, in the terminology used by Prof. Braumann in his presentation at the workshop such a replacement means a different *type of stochasticity.* Immigration from outside and changes of parameters inside a population is not the same. See in particular [3] for a more specific comparison, and e.g. [9] for a related problem of statistical inference. Prof. Braumann's presentation made it explicit that we usually have to pay a real price for replacing one type of stochasticity by another one, and this seems to be the case here as well.

This is one danger, but there is another one for RDBPs. If we model immigration we should model at the same time stochasticity which stems from the origin of immigrants. But then computability becomes a real problem.

To explain this, look again at the results of [7], and here specifically at Theorems 7.1 and 7.4. If the number of immigrants in a population becomes important then this may have a considerable impact on natality rates, and on productivity. In our model m and r would change, and it is good to see that the global model we proposed can cope with this. However, people with a different upbringing and origins also typically have a different consumption behaviour, hence implying in general a different claim distribution F. If we suppose that F can be suitably parametrised by m and r then (7.26) should now read

$$\Psi_{r,m}(dr, dm) := \frac{\partial m F^{(r,m)}(\tau[r,m])}{\partial r} dr + \frac{\partial m F^{(r,m)}(\tau[r,m])}{\partial m} dm$$

However, it is in practice not clear at all, whether such a parametrisation is meaningful. And then, why should $F^{(r,m)}(x)$ allow for a density? How should we then compute $\tau[r,m]$?

We conclude that those politicians who say that they understand the good and bad effects of immigration policies on social security and pensions must be creative mathematicians. Even for RDBPs which are still relatively simple models for societies compared to what societies seem to be in real life, a rigorous study of the influence of immigration is not evident at all.

References

1. Afanasyev, V.I., Geiger, J., Kersting, G., Vatutin, V.A.: Criticality in branching processes in random environment. Ann. Prob. **33**(2), 645–673 (2005)
2. Barbour, A.D., Hamza, K., Kaspi, H., Klebaner, F.C.: Escape from the boundary in Markov population processes. Adv. Appl. Probab. **47**(4) 1190–1211 (2015)
3. Braumann, C.: Environmental versus demographic stochasticity in population growth. In: González, M., del Puerto, I., Martínez, R., Molina, M., Mota, M., Ramos, A. (Eds.), Lecture Notes in Statistics - Proceedings, vol. 197, pp. 37–52. Springer, Berlin (2010)
4. Bruss, F.T.: A counterpart of the Borel-Cantelli Lemma. J. Appl. Prob. **17**, 1094–1101 (1980)
5. Bruss, F.T.: A note on extinction criteria for bisexual Galton-Watson processes. J. Appl. Prob. **21**, 915–919 (1984)
6. Bruss, F.T.: Grenzen einer jeden Gesellschaft (in German; English summary). In: Jahresb. der Deut. Math.-Ver., vol. 116, pp. 137–152. Springer, Berlin (2014)
7. Bruss, F.T., Duerinckx, M.: Resource dependent branching processes and the envelope of societies. Ann. Appl. Probab. **25**(1), 324–372 (2015)
8. Bruss, F.T., Robertson, J.B.: 'Wald's Lemma' for sums of order statistics of i.i.d. random variables. Adv. Appl. Probab. **23**, 612–623 (1991)
9. Bruss, F.T., Slavtchova-Bojkova, M.: On waiting times to populate an environment and a question of statistical inference. J. Appl. Probab. **36**, 261–267 (1999)
10. Daley, D.J., Hull, D.M., Taylor, J.M.: Bisexual Galton-Watson branching processes with superadditive mating functions. J. Appl. Probab. **23**, 585–600 (1986)
11. Haccou, P., Jagers, P., Vatutin, V.A.: Branching Processes - Variation, Growth and Extinction of Populations. Cambridge University Press, Cambridge (2005)
12. Hull, D.M.: A necessary condition for extinction in those bisexual Galton-Watson branching processes governed by superadditive mating function. J. Appl. Prob. **19**(4), 847–850 (1982)
13. Ispány, M.: Some asymptotic results for strongly critical branching processes with immigration in varying environment, Chapter 5. In: del Puerto, I., González, M., Gutiérrez, C., Martínez, R., Minuesa, C., Molina, M., Mota, M., Ramos, A. (eds.) Lecture Notes in Statistics - Proceedings. Springer, Berlin (2016)
14. Ispány, M., Pap, G., van Zuijlen, M.: Fluctuation limit of branching processes with immigration and estimation of the means. Adv. Appl. Probab. **37**, 523–538 (2005)
15. Jagers, P., Klebaner, F.C.: Population-size-dependent and age-dependent branching processes. Stoch. Process. Appl. **87**, 235–254 (2000)
16. Jagers, P., Klebaner, F.C.: Population-size-dependent, age-structured branching processes linger around their carrying capacity. J. Appl. Prob. **48A**, 249–260 (2011)
17. Klebaner, F.C.: On population-size dependent branching processes. Adv. Appl. Probab. **16**, 30–55 (1984)
18. Molina, M.: Two-sex branching process literature. In: González, M., del Puerto, I., Martínez, R., Molina, M., Mota, M., Ramos, A. (eds.) Lecture Notes in Statistics - Proceedings, vol. 197, pp. 279–291. Springer, Berlin (2010)
19. Wajnberg, A.: Le théorème de Bruss-Duerinckx. FNRS-News, Fond National de la Recherche Scientifique **6**, 21–12 (2014)
20. Wei, C.Z., Winnicki, J.: Some asymptotic results for the branching process with immigration. Stoch. Process. Appl. **31**, 261–282 (1989)
21. Yanev, N.M.: Conditions for degeneracy of φ-branching processes with random φ. Theory Probab. Appl. **20**, 421–428 (1976)

Chapter 8
From Size to Age and Type Structure Dependent Branching: A First Step to Sexual Reproduction in General Population Processes

Peter Jagers and Fima C. Klebaner

Mathematics Subject Classification (2000): 60J80, 92D25

8.1 From Population Size Dependence to Age Structure Effects

Classical branching processes, even the most general, share the property that individuals are supposed to multiply independently of one another, at least given some environment that in its turn is supposed to be unaffected by the population. Only more recently have birth-and-death and branching processes been considered which allow individual reproduction to be influenced by population size. The first results, due to Klebaner [14], deal with Galton-Watson processes. Work on general, age-structured processes and habitats with a threshold, a so called *carrying capacity*, came only decades later, cf. [5, 11, 13, 15, 16, 21], inspired by deterministic population dynamics [6–8, 22].

P. Jagers (✉)
Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, Gothenburg, Sweden
e-mail: jagers@chalmers.se

F.C. Klebaner
School of Mathematical Sciences, Monash University, Clayton, VIC, Australia
e-mail: fima.klebaner@monash.edu

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_8

The general formulation of population-size-dependence runs as follows. Any individual of age a in a population of size z in a habitat of carrying capacity K is characterised by three entities, which we refer to as *population parameters*:

- a birth intensity $b_z^K(a)$,
- a death intensity $h_z^K(a)$, and
- a probability distribution for the number of children at death at age a, $\{p_z^K(a,k), k = 0, 1, 2, \ldots\}$.

The birth intensity is supposed to catch the reproduction of higher animals and plants, during life, whereas the number of children at death (splitting) reflects cell division or particle split. Births during mother's life are assumed to occur one by one, and the distribution of the number of children at splitting should have bounded mean $m_z^K(a)$ and variance. Further, the processes are supposed to be monotonous in the sense that if we consider two classical branching process with the same starting condition and fixed parameters pertaining to "sizes" $z \leq z'$, the former population should be larger than or equal to the latter, in distribution, at any time. At population size K, i.e. for parameters b_K^K, h_K^K, etc. the population size-dependent process should pass from super- to sub-criticality, for precise formulations cf. [13]. In this section, we recapitulate results there and in [11, 12].

Let Z_t denote the population size at time $t \geq 0$ and write $T_a := \inf\{t; Z_t \geq aK\}$, where $0 < a < 1$ and $Z_0 = z$ is presumed (much) smaller than aK. Let Y_t be the (not population-size-dependent) process with parameters b_{aK}^K, h_{aK}^K, etc, i.e. frozen at population size aK, and $Y_0 = z$. Then, by the assumed monotonicity in parameters, $Z_t \geq Y_t$ on $\{T_a \geq t\}$ a. s.

Write T for the time to extinction, $\inf\{t; Z_t = 0\}$, so that the probability of early extinction can be interpreted as the probability of never attaining aK,

$$\mathbb{P}(T < T_a) = \mathbb{P}(\sup Z_t < aK) \leq \mathbb{P}(\sup Y_t < aK) = \mathbb{P}(Y_t \to 0) = q(aK)^z,$$

if $q(y)$ denotes the classical extinction probability of a process from one ancestor with parameters frozen at population size y. However, by coupling our process to the not-size-dependent branching process with parameters frozen at the initial size, much as customary in epidemic spread process, [3, 4, 19], we should be able to do better during an initial phase of length $o(\sqrt{K})$ or even $O(K^{2/3})$.

Informally, the coupling process can be described the following way, in the case where the life span distribution is unaffected by population size, and there is no splitting: Let $\tau(1), \tau(2), \ldots$ be the successive birth moments in the process with frozen parameters and denote the age of the mother in question at those times by $\alpha_1, \alpha_2, \ldots$. As before, write $Z_0 = z$. To recover the size-dependent population, then retain the ith birth with probability $\min(1, b_{Z_{\tau(i)}}^K(\alpha_i)/b_z^K(\alpha_i))$, provided none of her (grand)mothers has been deleted.

Also if the population escapes early extinction, it will satisfy $Z_t \geq Y_t$ as long as $Z_t \leq aK$. Since Y_t grows exponentially when not dying out, the Z process must then attain the level aK within a time of order $\log K$. Once up in a band around the carrying capacity, the population size can be expected to remain there for a time of length $O(e^{cK})$, for some $c > 0$ as $K \to \infty$, [13]. In special cases constants can be calculated. Thus [11, 15], for binary Galton-Watson process with Michaelis-Menten style reproduction, (i.e. the probability in the $n + 1$th generation of two children being = $K/(K + Z_n)$), the subsistence time will be $\sim e^{f(a)K}$, where

$$f(a) = \frac{a(1-a)^2}{8(1+a)}$$

for any $0 < a < 1$.

During the long sojourn around the carrying capacity, the age distribution will settle around a stable McKendrick-von Foerster form [10, 11], known from deterministic population dynamics. When ultimately the population embarks on its journey to extinction, i.e. passes the level aK never to reach it again, its remaining life will again be short, logarithmic in K, [11].

From size-dependence the step seems short, almost of notational character, to age structure dependence, as done in [11]. Indeed, let A_t denote the whole vector of ages at time $t \geq 0$, so that $Z_t = (1, A_t) = |A_t|$ and index parameters with these vectors instead of with just population size, b_A, h_A etc., $A = (a_1, a_2 \dots a_z)$. If the process is still supercritical below the carrying capacity, *strongly critical* at it, i.e. $|A| = K \Rightarrow b_A^K + h_A^K(m_A^K - 1) = 0$, and subcritical when large (and further technical conditions, cf, [11]), results for size dependence continue to hold.

Alas, the generalisation is not as far-reaching, as it might first seem. It opens up for dependence upon other population features, like volume or DNA mass dependence, but it is difficult to see realistic dependencies upon the age distribution per se. Still the situation is very much like that of population size dependence. It becomes radically different in multi-type cases, like sexual reproduction where only one type of individuals can reproduce, but in a manner that is heavily dependent upon the prevalence of the other type, in fertile ages.

8.2 Process Structure

Hence, we take a further step and lay the foundations for mathematical analysis of multi-type processes with population structure dependence. The individuals in the population are now characterised by their types and ages, to be written $s_1, s_2, \dots, s_z$, where $s_j = (i_j, a_j) \in \mathcal{S} = \mathbb{T} \times \mathbb{R}^+$, and z is the population size. The type space $\mathbb{T}$ is finite, say consisting of k types, and the state of the process S is taken to be the array $s_1, s_2, \dots, s_z$ or equivalently the counting measure

$$S(B) = \sum_{j=1}^{z} \delta_{s_j}(B), B \subset \mathcal{S}. \tag{8.1}$$

For a test function $f : \mathcal{S} \to \mathbb{R}$, hence,

$$(f,S) = \int f(s)S(ds) = \sum_{i=1}^{z} f(s_j). \tag{8.2}$$

Population evolution is given by a measure-valued process $\{S_t; t \geq 0\}$, reporting the number of individuals, their ages, and their types at any time $t \geq 0$. The randomness comes through random events of birth and death. These are described, as usual, by their intensities, which besides dependence on the individual's own age and type may be influenced by the whole population structure, i.e. by the other individuals' ages and types.

The process develops in a simple manner: while there are no deaths or births, it changes only by ageing. When an individual dies, its point mass disappears and an offspring number of point masses at zero age appears. Similarly, when she gives birth during life a point mass appears at the origin. Type at birth is determined by the type distribution, for example sex may be obtained by independent Bernoulli trials with probability p for becoming female and $q = 1 - p$ for male. More generally we write $\mathbf{p} = (p_1, p_2, \ldots, p_k)$ for the probabilities that a type variable I takes the values $1, 2 \ldots k$. These are also allowed to depend on the mother's age and type, s, and the whole population structure S; we write $\mathbf{p}_S(s)$. The reason to include reproduction both at death (splitting) and during life time is to cover cases corresponding to Galton-Watson as well as birth-and-death processes. To avoid unnecessary technicalities we assume throughout that all quantities below (like derivatives, expectations, etc.) exist and are well defined

Theorem 8.1 *For any bounded C^1 function F on $\mathbb{R}^+$ and function f on $\mathcal{S}$, which is C^1 in its last argument,*

$$\lim_{t \to 0} \frac{1}{t} \mathbb{E}_S[F((f,S_t)) - F((f,S))] = \mathcal{G}F((f,S)), \tag{8.3}$$

where

$$\mathcal{G}F((f,S)) = F'((f,S))(f',S) + \sum_{j=1}^{z} b_S(s_j)[\mathbb{E}_{p_S(s_j)} F(f(I,0) + (f,S)) - F((f,S))]$$
$$+ \sum_{j=1}^{z} h_S(s_j)\mathbb{E}[F(\sum_{l=1}^{k} f(l,0)Y_S^l(s_j) + (f,S) - f(s_j)) - F((f,S))].$$

Here, f' is the derivative in the second co-ordinate a, and $Y_S^l(i,a)$ is the number of l-type offspring of an i-type individual dying at age a, the population composition then being S.

In what follows, $m^l_S(s) = \mathbb{E}_S Y^l(s)$, $l = 1, 2 \dots, k$, denotes the means and

$$\Gamma = (\gamma_{l_1,l_2}) = \Big(\mathbb{E}_S Y^{l_1}(s) Y^{l_2}(s)\Big)_{l_1,l_2}$$

the matrix of second moments of offspring at death distributions. $\mathcal{G}$ defines a generator of a measure-valued multi-type branching process. This is Dynkin's formula:

Theorem 8.2 *For any bounded C^1 function F on $\mathbb{R}^+$ and function f on $\mathcal{S}$, which is C^1 in its last argument,*

$$F((f, S_t)) = F((f, S_0)) + \int_0^t \mathcal{G}F((f, S_u))du + M^{F,f}_t, \tag{8.4}$$

where $M^{F,f}_t$ is a zero mean local martingale with predictable quadratic variation

$$\Big\langle M^{F,f}_t, M^{F,f}_t \Big\rangle = \int_0^t \Big(\mathcal{G}F^2((f, S_u)) - 2F((f, S_u))\mathcal{G}F((f, S_u))\Big)du, \tag{8.5}$$

and variance

$$\mathbb{E}_S (M^{F,f}_t)^2 = \mathbb{E}_S \Big\langle M^{F,f}_t, M^{F,f}_t \Big\rangle. \tag{8.6}$$

The derivation of the above is the same as in [12].

Taking the identity map, $F^{id}(u) = u$, as F above, results in the linear operator L_S obtained by $\mathcal{G}F^{id}$.

Theorem 8.3 *For a function f on $\mathcal{S}$, which is C^1 in its last argument,*

$$(f, S_t) = (f, S_0) + \int_0^t (L_{S_u} f, S_u)du + M^f_t, \tag{8.7}$$

where (note that p^l_S and m^l_S are functions on $\mathcal{S}$)

$$L_S f = f' - h_S f + \Big(\sum_{l=1}^k f(l, 0) p^l_S\Big) b_S + \Big(\sum_{l=1}^k f(l, 0) m^l_S\Big) h_S, \tag{8.8}$$

and M^f_t is a local square integrable martingale with the sharp bracket given by

$$\big\langle M^f, M^f \big\rangle_t = \int_0^t \Big((\sum_{l=1}^k f^2(l, 0) p^l_{S_u}) b_{S_u} + \Big(\sum_{l_1=1}^k \sum_{l_2=1}^k f(l_1, 0) f(l_2, 0) \gamma_{l_1 l_2}\Big) h_{S_u}$$
$$-2 \sum_{l=1}^k f(l, 0) m^l_{S_u} h_{S_u} f + h_{S_u} f^2, S_u\Big) du.$$

The special choice of f such that $f(s) = 1$ identically yields the equation for the evolution in time t of the total population size, whereas taking f to be $\delta_i(s) = 1$ as soon as the first coordinate of s is i and zero otherwise gives the evolution of the subpopulation of type i individuals.

Thus, the equation for the total size $Z_t = (1, S_t)$ reduces to

$$Z_t = Z_0 + \int_0^t (b_{S_u} + h_{S_u}(m_{S_u} - 1), S_u)du + M_t^1, \tag{8.9}$$

where $m_S = \sum_{l=1}^k m_S^l$ is the mean of the total number of offspring at death and

$$\langle M^Z, M^Z \rangle_t = \int_0^t \Big(b_{S_u} + \Big(\sum_{l_1=1}^k \sum_{l_2=1}^k \gamma_{l_1 l_2} \Big) h_{S_u} - 2m_{S_u} h_{S_u} + h_{S_u}, S_u \Big) du. \tag{8.10}$$

For the type i subpopulation $Z_t^i = (\delta_i, S_t)$,

$$Z_t^i = Z_0^i + \int_0^t (-h_{S_u}\delta_i + p_{S_u}^i b_{S_u} + m_{S_u}^i h_{S_u}, S_u)du + M_t^i, \tag{8.11}$$

$$\langle M^i, M^i \rangle_t = \int_0^t \Big(p_{S_u}^i b_{S_u} + \gamma_{ii} h_{S_u} - 2m_{S_u}^i h_{S_u} \delta_i + h_{S_u}\delta_i, S_u \Big) du. \tag{8.12}$$

8.3 Processes with Carrying Capacity

As mentioned, the idea of a carrying capacity K is that population parameters should stabilise for large K in a manner that makes population size fluctuate around it, if it ever came close to it, cf. [10, 11, 13] in the single-type case. This is made precise in the assumptions about demographic smoothness to follow. We thus study populations, which have a carrying capacity K, and superscript corresponding entities with a K, thought of as large. Since we are interested in their behaviour around this carrying capacity, we assume that they start close to it, and study their type and age composition while they prevail there, proving a Law of Large Numbers corresponding to the single-type case of [10].

Thus, consider measures of the form

$$\bar{S}^K = \frac{1}{K} S^K, \tag{8.13}$$

which are of course no longer counting measures on $\mathcal{S}$ but still have finite support. In the limit as $K \to \infty$, the age distribution turns out absolutely continuous, whereas the type distribution remains discrete on the given space with k elements.

A population process will be said to be *demographically smoothly density dependent*, or for short just demographically smooth, if:

C0. The model parameters, b^K, h^K, m^K's, and Γ^K's are uniformly bounded.

C1. They are also normed uniformly Lipschitz in the following sense: there is a $C > 0$ such that for all $u \in \mathcal{S} = \mathbb{T} \times \mathbb{R}^+$ and K, $\rho(\mu, \nu)$ denoting the Levy-Prokhorov distance between measures μ and ν on $\mathcal{S}$ (rather on corresponding σ-algebra),

- $|b^K_\mu(u) - b^K_\nu(u)| \leq C\rho(\mu/K, \nu/K)$,
- $|h^K_\mu(u) - h^K_\nu(u)| \leq C\rho(\mu/K, \nu/K)$,
- $|m^{l,K}_\mu(u) - m^{l,K}_\nu(u)| \leq C\rho(\mu/K, \nu/K), l = 1, 2 \ldots, k$,
- $|p^{l,K}_\mu(u) - p^{l,K}_\nu(u)| \leq C\rho(\mu/K, \nu/K), l = 1, 2 \ldots, k$.

C2. If the measures μ^K on $\mathcal{S}$ are such that $\bar{\mu}^K = \mu^K/K$ converges weakly to some $\bar{\mu}$, as $K \to \infty$, then the limit $\lim b^K_{\mu^K} = b^\infty_{\bar{\mu}}$ exists, the same applying to limits $\lim h^K_{\mu^K} = h^\infty_{\bar{\mu}}$ and all $\lim m^{l,K}_{\mu^K} = m^{l,\infty}_{\bar{\mu}}$ as well as $\lim p^{l,K}_{\mu^K} = p^{l,\infty}_{\bar{\mu}}$.

C3. $\bar{S}^K_0 \Rightarrow \bar{S}^\infty_0$, and $\sup_K |\bar{S}^K_0| < \infty$, i.e. the process *stabilises initially*.

Remark As an illustration, the dependence of parameters upon the type and age composition could have the form of a dependence on $\bar{S}^K$, e.g. on $(\delta_i\phi, \bar{S}^K)$, where ϕ is a function of age only. For example, if $\phi(a) = I_{[a_1,a_2]}(a)$, we would obtain the density of i-individuals in an age interval like that of reproductive ages. For such dependencies assumption C2 is satisfied, if $b^K_\mu = b_{(\phi,\bar{\mu}^K)} \to b_{(\phi,\bar{\mu})}$ depends on its suffix in a continuous fashion—similarly for other parameters.

The following result gives the stable type and stable age distributions $\bar{S}^\infty_t$. It is obtained from (8.7) by dividing through by K and taking limits by using the smooth demography assumption. The martingale term vanishes (eg. by Doob's inequality) and we have the main result:

Theorem 8.4 *Under the condition of demographical smoothness, the measure-valued density process* $\{\frac{1}{K}S^K_t; t \geq 0\}$ *converges weakly, as* $K \to \infty$, *to a process* $\{\bar{S}^\infty_t; t \geq 0\}$, *which solves the following equation*

$$(f, S_t) = (f, S_0) + \int_0^t (L_{S_u} f, S_u) du \tag{8.14}$$

for test functions f. *Here* L_S *is given in* (8.8), *with parameters being the limiting quantities in the smooth demographics.*

Next, we make Eq. (8.14) explicit by using disintegration of measures S_t ($\bar{S}^\infty_t$), cf. [20], e.g. The resulting explicit form is helpful in, for example, the case of sexual reproduction.

Since $\mathbb{T}$ is a discrete and indeed finite set, we can write

$$\pi^i = S(\{i\} \times \mathbb{R}^+), \quad A^i(da) = \frac{1}{\pi^i} S(\{i\} \times da),$$

provided π^i is positive. In our context π^i is the proportion of type i individuals and A^i is the age distribution of type i (conditioned on the individual being of type i). Hence, $\bar{S}_t^\infty$ gives the limiting proportions of types and the age distribution of each type. If $\pi^i = 0$, then A^i is not defined, which is of no consequence as it will disappear from the integral below. For test functions f on $\mathcal{S} = \mathbb{T} \times \mathbb{R}^+$,

$$(f, S) = \int_{\mathcal{S}} f(u)S(du) = \sum_{i=1}^{k} \pi^i \int_{\mathbb{R}+} f(i,a)A^i(da),$$

so that the limit equation (8.14) for $\bar{S}_t^\infty$ can be written explicitly

$$\sum_{i=1}^{k} \pi_t^i \int_{\mathbb{R}+} f(i,a)A_t^i(da) = \sum_{i=1}^{k} \pi_0^i \int_{\mathbb{R}+} f(i,a)A_0^i(da) + \sum_{i=1}^{k} \int_0^t \pi_u^i \int_{\mathbb{R}+} g_{S_u}(i,a)A_u^i(da)du, \tag{8.15}$$

where

$$\begin{aligned} g_S(i,a) = L_S f(i,a) &= f'(i,a) - h_S(i,a)f(i,a) + \Big(\sum_{l=1}^{k} f(l,0)p_S^l(i,a)\Big)b_S(i,a) \\ &\quad + \Big(\sum_{l=1}^{k} f(l,0)m_S^l(i,a)\Big)h_S(i,a), \end{aligned} \tag{8.16}$$

omitting the superscript '∞' in the limiting quantities, as specified in C2. From this, we can find equations both for the type distribution and for the age distribution of every type. This is done in detail in the next section for the case of sexual reproduction.

8.4 Sexual Reproduction

We specialise to the case of there being two types, females, and males; only females reproducing, but with a fertility, that can be influenced by the availability of males. The idea is that this will render the study of sexual reproduction possible also in age-structured models, circumventing the notorious question of how to model couple formation in structured populations.

In the unstructured Galton-Watson case, sexual reproduction has been extensively studied for almost half a century, cf. relevant sections in [9] and the sequel of papers from the Extremadura school, [17], and others. The success hinges upon mating occurring in the one and only generation in existence at any given moment, and more realistic population models, like general branching processes, allowing co-existence of individuals from several generations, have resisted attempts at generalisation to sexual reproduction. The same applies to age-structured deterministic population dynamics, which also remains asexual, cf. [7, 8], and even to demographic theory, which is clonal in spite of it study object being human populations (this probably warranted by the large number context of demography). For a simple but elegant approach to reproduction with mating in the birth and death case cf. [1, 2], generalised in [18].

We thus specialise the framework of the preceding section by choosing the number of types $k = 2$, and write the type space $\mathbb{T} = \{1, 2\}$. Since only females can give birth, we may simplify notation further by writing b_S as a function of age alone for the female birth rate and similarly for the probability p_S of the child being female. The probability of a male child is $q_S = 1 - p_S$. Similarly, the mean number of females born at death is m_S^1 and males m_S^2, and the hazard rate for females is h_S^1 and for males h_S^2, again functions of age, i.e. $h_S^i(a) = h_S(i, a), i = 1, 2$. Omitting the superscript ∞ in the limiting quantities as specified in C2, (see also the Remark) we see that the operator L_S, in (8.8) and (8.16), takes the form

$$\begin{aligned} L_S f(1,a) &= f'(1,a) - h_S^1(a) f(1,a) + \Big(f(1,0) p_S(a) + f(2,0) q_S(a) \Big) b_S(a) \\ &\qquad + \Big(f(1,0) m_S^1(a) + f(2,0) m_S^2(a) \Big) h_S^1(a), \\ L_S f(2,a) &= f'(2,a) - h_S^2(a) f(2,a). \end{aligned} \tag{8.17}$$

If we introduce test functions of age, ϕ and ψ, and write first $f(1,a) = \phi(a)$ and $f(2,a) = 0$, and then $f(1,a) = 0$ and $f(1,a) = \psi(a)$, the limit equation (8.14), (8.15) becomes as follows. Here it is easier to work with $S_t^i = \pi_t^i A_t^i$.

$$(\phi, S_t^1) = (\phi, S_0^1) + \int_0^t \Big(\phi' - h^1_{S_u} \phi + \phi(0)(p_{S_u} b_{S_u} + m^1_{S_u} h^1_{S_u}), S_u^1 \Big) du, \tag{8.18}$$

$$(\psi, S_t^2) = (\psi, S_0^2) + \int_0^t \Big[\Big(\psi' - h^2_{S_u} \psi, S_u^2 \Big) + \Big(\psi(0) q_{S_u} b_{S_u} + \psi(0) m^2_{S_u} h^1_{S_u}, S_u^1 \Big) \Big] du.$$

By integration by parts and looking at the adjoint operators, one can show that the densities $s^i(t,x)$ on $(0,\infty)$ of S_t^i exist ($i = 1, 2$), and that these equations are the weak form of the pde's, [10].

$$\frac{\partial s^1}{\partial t} + \frac{\partial s^1}{\partial x} = -h_{S_t}(s^1), \quad s^1(t,0) = \int_0^\infty (p_{S_t} b_{S_t} + m^1_{S_t} h^1_{S_t})(y) s^1(t,y) dy, \tag{8.19}$$

$$\frac{\partial s^2}{\partial t} + \frac{\partial s^2}{\partial x} = -h_{S_t}(s^2), \quad s^2(t,0) = \int_0^\infty (q_{S_t} b_{S_t} + m^2_{S_t} h^1_{S_t})(y) s^1(t,y) dy.$$

Analysis of these two-dimensional generalised McKendrick-von Foerster equations will be given elsewhere.

Further, we derive equations for the types. If $f = \delta^1$,

$$L_S f(1,a) = h^1_S(a)(m^1_S(a) - 1) + p_S(a) b_S(a), \;\; L_S f(2,a) = 0,$$

and for $f = \delta^2$,

$$L_S f(1,a) = q_S(a) b_S(a) + m^2_S(a) h^1_S(a), \;\; L_S f(2,a) = -h^2_S(a).$$

This implies the following equation for the limiting density of females $\bar{X}_t = \pi^1_t$, and males $\bar{Y}_t = \pi^2_t$,

$$\bar{X}_t = \bar{X}_0 + \int_0^t \bar{X}_u \int_{\mathbb{R}+} \Big[h^1_{S_u}(a)(m^1_{S_u}(a) - 1) + p_{S_u}(a) b_{S_u}(a) \Big] A^1_u(da) du \tag{8.20}$$

$$\bar{Y}_t = \bar{Y}_0 + \int_0^t \Big[\bar{X}_u \int_{\mathbb{R}+} (q_{S_u}(a) b_{S_u}(a) + m^2_{S_u}(a) h^1_{S_u}(a)) A^1_u(da) - \bar{Y}_u \int_{\mathbb{R}+} h^2_{S_u}(a) A^2_u(da) \Big] du,$$

which can be written as coupled ordinary differential equations,

$$\dot{\bar{X}}_t = \bar{X}_t (h^1_{S_t}(m^1_{S_t} - 1) + p_{S_t} b_{S_t}, A^1_t). \tag{8.21}$$

$$\dot{\bar{Y}}_t = \bar{X}_t (q_{S_t} b_{S_t} + m^2_{S_t} h^1_{S_t}, A^1_t) - \bar{Y}_t (h^2_{S_t}, A^2_t).$$

This last form shows how parameters of the model combine to influence the composition and evolution, and should be analysed for fixed points etc.. It can be viewed as a sort of generalised Lotka-Volterra equation, most obviously so in the particular case, where the demographic parameters are independent of the individual's age and depend upon S_t only through the number of males and females X_t, Y_t, or only Y_t.

8.5 Future Challenges

It may be suitable to factor female birth intensities into a probability of being fertilised, $\phi(x,y)$, a function of the numbers x and y of females and males present at each time point, whereas crowding effects are mirrored in population size dependence of the reproduction intensity of fertile females β_z and of the death intensities h_z,

$$b_{S_t} = \phi(X_t, Y_t)\beta_{Z_t}, m_{S_t} = \phi(X_t, Y_t)\mu_{Z_t}, h_{S_t} = h_{Z_t}.$$

Clearly $\phi(x,0)$ ought to be taken as 0 for all x. The case where ϕ is an increasing function of the number of males only, reflects promiscuity, whereas a situation where it is a function of the sex ratio Y_t/X_t comes closer to mating proper. In any case, the probability of fertilisation should be taken as an increasing function of its second argument, maybe satisfying $\phi(x, 1) \geq \rho$ for some $\rho > 0$ and all $x \geq 1$. (This rules out the case of strict and complete monogamy.) In the case of a carrying capacity K the parameters β_z^K, μ_z^K, and h_z^K define supercritical populations for $z < K$ turning subcritical as $z > K$, in the sense discussed in [13].

We have outlined an approach to type and age structured population dynamics with a possible application to sexual reproduction in general stochastic population models. It raises many questions, some of which have been noted in the text. Others concern extinction probabilities, growth rates, prevalence time, and sex ratios and other compositional matters during prevalence, and of course about differences to and similarities with clonally reproducing populations, as well as between different fertilisation schemes. Investigations about Galton-Watson processes with mating indicate that clear general assertions about the initial phase, survival chances and a stage of exponential growth, may be hard cf. [17] or Alsmeyer's sections in [9], whereas the approach in [10–13] may give some hope for a description of the lingering phase. We intend to pursue this in forthcoming more general work.

Acknowledgements The first author is grateful to the Royal Society of Sciences and Arts in Gothenburg for supporting his participation in the Workshop, and thereby also this survey of his lecture. The work is part of a wider project, supported by the Australian Research Council under grant DP150103588.

The article has benefitted greatly from the close reading by an anonymous referee.

References

1. Asmussen, S.: On some two-sex population models. Ann. Probab. **4**, 727–744 (1980)
2. Asmussen, S., Hering, H.: Branching Processes. Birkhäuser, Boston (1983)
3. Ball, F.G.: The threshold behaviour of epidemic models. J. Appl. Probab. **20**, 227–241 (1983)
4. Ball, F.G., Donnelly, P.: Strong approximations for epidemic models. Stoch. Process. Appl. **55**, 1–21 (1995)
5. Champagnat, N., Ferriere, R., Méléard, S.: From individual stochastic processes to macroscopic models in adaptive evolution. Stoch. Models **24**, 2–44 (2008)
6. Cushing, J.M.: Existence and stability of equilibria in age-structured population dynamics. J. Math. Biol. **20**, 259–276 (1984)
7. Diekmann, O., Gyllenberg, M., Metz, J.: On the formulation and analysis of general deterministic structured population models I. Linear theory. J. Math. Biol. **36**, 349–388 (1998)
8. Diekmann, O., Gyllenberg, M., Metz, J., et al.: On the formulation and analysis of general deterministic structured population models II. Nonlinear Theory. J. Math. Biol. **43**, 157–189 (2001)
9. Haccou, P., Jagers, P., Vatutin, V.A.: Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press, Cambridge (2005)
10. Hamza, K., Jagers, P., Klebaner, F.K.: The age structure of population-dependent general branching processes in environments with a high carrying capacity. Proc. Steklov Inst. Math. **282**, 90–105 (2013)
11. Hamza, K., Jagers, P., Klebaner, F.K.: On the Establishment, Persistence and Inevitable Extinction of Populations. J. Math. Biol. (2015). doi:10.1007/s00285-015-0903-2

12. Jagers, P., Klebaner, F.C.: Population-size-dependent and age-dependent branching processes. Stoch. Process. Appl. **87**, 235–254 (2000)
13. Jagers, P., Klebaner, F.C.: Population-size-dependent, age-structured branching processes linger around their carrying capacity. J. Appl. Probab. **48A**, 249–260 (2011)
14. Klebaner, F.C.: On population-size-dependent branching processes. Adv. Appl. Probab. **16**, 30–55 (1984)
15. Klebaner, F.C., Sagitov, S., Vatutin, V.A., Haccou, P., Jagers, P.: Stochasticity in the adaptive dynamics of evolution: the bare bones. J. Biol. Dyn. **5**, 147–162 (2011)
16. Méléard, S., Tran, V.C.: Trait substitution sequence process and canonical equation for age-structured populations. J. Math. Biol. **58**, 881–921 (2009)
17. Molina, M.: Two-sex branching process literature. In: González, M., et al. (eds.) Workshop on Branching Processes and Their Applications. Springer Lecture Notes in Statistics, vol. 197, pp. 279–293. Springer, Berlin (2010)
18. Molina, M., Yanev, N.: Continuous time bisexual branching processes. C. R. Acad. Bulg. Sci. **56**, 5–10 (2003)
19. Reinert, G., Barbour, A.: Approximating the epidemic curve. Electron. J. Probab. **18**, Article 54 (2013). doi:10.1214/EJP.v18-2557
20. Schwartz, L.: Lectures on Disintegration of Measures. Tata Institute of Fundamental Research, Bombay (1976)
21. Tran, V.C.: Large population limit and time behaviour of a stochastic particle model describing an age-structured population. ESAIM Probab. Stat. **12**, 345–386 (2008)
22. Webb, G.F.: Theory of Nonlinear Age-Dependent Population Dynamics. Dekker, New York (1985)

Part V
Age-Dependent Branching Models

Chapter 9
Supercritical Sevastyanov Branching Processes with Non-homogeneous Poisson Immigration

Ollivier Hyrien, Kosto V. Mitov, and Nikolay M. Yanev

Mathematics Subject Classification (2000): 60J80, 60F05, 60J85, 62P10

9.1 Introduction

Sevastyanov [26] introduced branching processes with immigration and studied their properties in the Markov case when immigration occurs in accordance with a time-homogeneous Poisson process. Extensions to the Bellman–Harris process with time-homogeneous immigration were subsequently considered by Jagers [14], Pakes [23], Radcliffe [25], Pakes and Kaplan [24], and Kaplan and Pakes [16]. Mitov and Yanev [18–20] investigated Bellman–Harris processes with state-dependent immigration (see also Chap. 3 in Ahsanullah and Yanev [1]).

The Sevastyanov process is an extension of the Bellman–Harris process, which allows the lifespan and the offspring of the process to be dependent [27–29]. This extension has been supported by analysis of cell kinetics data [4, 10–12].

O. Hyrien
Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA
e-mail: Ollivier_Hyrien@urmc.rochester.edu

K.V. Mitov (✉)
Faculty of Aviation, National Military University "Vasil Levski", 5856 D. Mitropolia, Pleven, Bulgaria
e-mail: kmitov@yahoo.com

N.M. Yanev
Department of Probability and Statistics, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
e-mail: yanev@math.bas.bg

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_9

Comprehensive reviews of branching processes, including their applications to biology, can be found in the following monographs: Harris [7], Sevastyanov [29], Mode [22], Athreya and Ney [3], Jagers [15], Asmussen and Hering [2], Yakovlev and Yanev [32], Kimmel and Axelrod [17], Haccou et al. [6] and Ahsanullah and Yanev [1]. For specific discussions on applications of Markov and Bellman–Harris branching processes with non-homogeneous Poisson immigration to modeling cell kinetics, the reader is referred to Yakovlev and Yanev [33, 34] and Hyrien and Yanev [8, 9]. See also Hyrien et al. [13] for an application of a two-type Bellman–Harris process with immigration to study the dynamics of immature (BFU-E) and mature (CFUE) erythroid progenitors following sublethal total body irradiation.

The model under consideration herein is developed based on Sevastyanov [27–29] age-dependent branching process and, in addition, allows an immigration component at times described by a time-nonhomogeneous Poisson process. In fact the present paper offers a continuation of earlier investigations by Mitov and Yanev [21] who focused on the critical case. Here, we study the supercritical case. Note that this case was investigated by Yanev [30, 31] when immigration is time-homogeneous.

The paper is organized as follows. Section 9.2 describes the biological background and motivation. This will provide the intuition used to construct the more general model formulated in Sect. 9.3. The basic equations for the probability generating functions and the moments are also presented in Sect. 9.3. The asymptotic behavior for the means, variances and covariances of the supercritical process with immigration are investigated in Sect. 9.4 for several classes of immigration rates. In Sect. 9.5 we prove two sets of limit results. The first one (Theorem 9.1) generalizes classical results obtained by Sevastyanov [26], while the second one (Theorems 9.2 and 9.3) describes novel behaviors that arise from the non-homogeneity of the immigration process. Theorem 9.2 can be interpreted as a LLN and Theorem 9.3 as a CLT.

9.2 Biological Motivation

The development and repair of tissues of the body is controlled by the processes of cell division, cell death, and cell differentiation. Because the outcome of the cell cycle is stochastic, age-dependent branching processes have been proposed to describe the dynamics of cell populations. To date, the Bellman–Harris process has remained the model of choice in such applications. It is flexible enough to accommodate multiple cell types, and it allows the possibility for cells to immigrate from an unobservable compartment via an immigration component [8, 33]. An important assumption made by this model is that the duration of the lifespan and the outcome of the cell cycle (that is, division, death, or differentiation) are independent. In other words, the conditional offspring distribution (here, the probability that the cell either divides, dies, or differentiates), given the lifespan of the cell, given the lifespan duration, does not change during the lifespan.

Several studies have shown that the duration of the cell lifespan and the outcome of the cell cycle are stochastically dependent. For example, studies that investigated the generation of oligodendrocytes from their cultured progenitor cells found that these cells self-renewed and differentiated into oligodendrocytes dissimilarly over time [4, 10, 11]. Other studies reported that lymphocytes were prone to cell death about 1 h after birth but divided only after about 12 h on average [12]. These findings are consistent with the assumptions of the Sevastyanov process.

A population in which cells divide, differentiate, die, or immigrate into the population could be described as follows. The process begins at time zero with N_0 cells. When modeling the development of tissues from the earliest embryonic stages it is reasonable to set $N_0 = 0$. New cells (immigrants) arrive in the population in accordance with a non-homogeneous Poisson process with instantaneous rate $r(t)$. These immigrants are assumed to be of age zero upon arrival in the population. When completing its lifespan, every cell either divides into two new cells or it exits the mitotic cycle to either differentiate or to die. Let p_2 denote the probability of division, and put $p_0 = 1 - p_2$ for the probability of exiting the mitotic cycle. The duration of the lifespan of any cell (referring here to either the time to division or the time to death/differentiation) is described by a non-negative random variable τ. Conditional on the cell dividing, τ has cumulative distribution function (c.d.f.) $G_2(t) = P\{\tau \le t|\xi = 2\}$ that satisfies $G_2(0) = 0$, whereas if it exits the mitotic cycle to either die or differentiate, it has a c.d.f. $G_0(x) = P\{\tau \le x|\xi = 0\}$ that satisfies $G_2(0) = 0$. Cells are assumed to evolve independently of each other.

The unconditional c.d.f of the duration of the lifespan, $G(t)$, is a mixture

$$G(t) = p_0 G_0(t) + p_2 G_2(t),$$

and the conditional distribution of ξ given τ is the posterior probability

$$P(\xi = x|\tau \ge t) = \frac{p_x(1 - G_x(t))}{1 - p_0 G_0(t) - p_2 G_2(t)},$$

for $x = 0$ or 2, such that the process belongs to the class of Sevastyanov processes when $G_0 \not\equiv G_2$. This example motivated the investigation of a class of Sevastyanov branching processes with non-homogeneous Poisson immigration defined in the next section. We consider a more general process to extend the scope of our work.

9.3 The Process and Its Equations

We define a process that begins with the immigration of a first group of cells, all of age zero, into the population at a random time S_1. Additional groups of cells, also of age zero, subsequently immigrate into the population at random time points S_k, $k = 1, 2\cdots$, ordered as $0 = S_0 < S_1 < S_2 < \cdots$. We assume that this sequence of random times forms a non-homogeneous Poisson process $\Pi(t)$ with instantaneous rate $r(t)$. The cumulative rate is $R(t) = \int_0^t r(u)du$. Let $U_i = S_i - S_{i-1}$, $i = 1, 2\cdots$, denote the inter-arrival times, such that $S_k = \sum_{i=1}^k U_i$, $k = 1, 2, \ldots$. Write I_k for the number of immigrants that enter the population at time S_k. We assume that $\{I_k\}_{k=1}^\infty$ is a sequence of independent and identically distributed r.v. with a p.g.f. $g(s) = E[s^{I_k}] = \sum_{i=0}^\infty g_i s^i$, $|s| \leq 1$. Let $\gamma = E[I_k] = \frac{dg(s)}{ds}|_{s=1}$ be the immigration mean and introduce the second factorial moment $\gamma_2 = \frac{d^2g(s)}{ds^2}|_{s=1} = E[I_k(I_k - 1)]$.

Upon arrival, every immigrant initiates an age-dependent branching process. This process describes the lifespan of every cell by a non-negative random variable η with c.d.f. $G(t) = P\{\eta \leq t\}, t \geq 0$. Upon completion of its lifespan, every cell produces a random number of offspring, say ξ, with conditional probability generating function (p.g.f.) given on the cell is of age u

$$h(u, s) = \sum_{k=0}^\infty p_k(u)s^k,\ h(u, 1) \equiv 1,\ |s| \leq 1.$$

We assume that every cell evolves independently of every other cell and that the collection of trees initiated by the immigrants are mutually independent.

The evolution of the process is governed by the joint distribution of (η, ξ). They need not be independent, in which case the above assumptions define a Sevastyanov branching process with immigration. It reduces to a Bellman–Harris process if the random variables η and ξ are independent; that is, if $h(u, s) \equiv h(s)$.

Let, for every $t \geq 0$, $Z(t)$ denote the number of cells at time t in the population for the process without immigration and started from a single cell at time $t = 0$. Sevastyanov [27] showed that the associated p.g.f. $F(t; s) = E[s^{Z(t)}|Z(0) = 1]$ satisfies the non-linear integral equation

$$F(t, s) = s(1 - G(t)) + \int_0^t h(u, F(t - u, s))dG(u) \tag{9.1}$$

with the initial condition $F(0, s) = s$ and under mild regularity conditions this equation admits a unique solution in the class of the p.g.f.

Define the moments of the offspring distribution:

$$a(u) = \left.\frac{\partial h(u, s)}{\partial s}\right|_{s=1}, \quad b(u) = \left.\frac{\partial^2 h(u, s)}{\partial s^2}\right|_{s=1}$$

$$a = \int_0^\infty a(u)dG(u), \quad \text{and} \quad b = \int_0^\infty b(u)dG(u).$$

and let $\mu = E[\eta] = \displaystyle\int_0^\infty x dG(x)$ denote the expected life-span of any cell. These characteristics are assumed to be finite. The Malthusian parameter α is the solution to the equation

$$\int_0^\infty e^{-\alpha x} a(x) dG(x) = 1.$$

This paper is concerned with the supercritical case $a > 1$, where α always exists and is strictly positive.

Define the moments

$$\begin{aligned}
&A(t) = E[Z(t)|Z(0) = 1],\ t \geq 0,\\
&B(t,\tau) = E[Z(t)Z(t+\tau)|Z(0) = 1],\ \ t \geq 0,\ \tau \geq 0,\\
&B(t) = E[Z(t)(Z(t)-1)|Z(0) = 1] = B(t,0) - A(t),\ t \geq 0,\\
&V(t) = \mathrm{Var}[Z(t)|Z(0) = 1] = B(t) + A(t) - A^2(t) = B(t,0) - A(t)^2,\ t \geq 0.
\end{aligned}$$

Throughout, we shall assume that $G(t)$ and $G_a^{(\alpha)}(t) = \int_0^t a(u)e^{-\alpha u} dG(u)$ are non-lattice distributions. Then, $A(t)$ satisfies the asymptotic approximation as $t \to \infty$ [29],

$$A(t) = Ae^{\alpha t}(1 + o(1)),\ \ A = \frac{\int_0^\infty e^{-\alpha t}(1 - G(t))dt}{\int_0^\infty x e^{-\alpha x} a(x) dG(x)}. \tag{9.2}$$

If, additionally, $\int_0^\infty b(u)e^{-2\alpha u} dG(u) < \infty$, then as $t \to \infty$,

$$B(t,\tau) = Be^{\alpha\tau + 2\alpha t}(1 + o(1)),\ \ B = \frac{A^2 \int_0^\infty b(x)e^{-2\alpha x} dG(x)}{1 - \int_0^\infty a(x)e^{-2\alpha x} dG(x)}. \tag{9.3}$$

Therefore, as $t \to \infty$, $B(t) = Be^{2\alpha t}(1 + o(1))$,

$$V(t) = B(t,0) - A(t)^2 = Ve^{2\alpha t}(1 + o(1)), \quad V = B - A^2. \tag{9.4}$$

Let $\{Y(t), t \geq 0\}$ denote the number of cells at time t as described by the process with immigration. Unless $G(t)$ is an exponential distribution, this process is non-Markov and can be studied by means of integral equations. It satisfies $Y(t) = 0$ as long as $\Pi(t) = 0$, and, when $\Pi(t) > 0$, it can be decomposed as

$$Y(t) = \sum_{k=1}^{\Pi(t)} \sum_{l=1}^{I_k} Z^{(k,l)}(t - S_k),$$

where $Z^{(k,l)}(t)$, $l = 1, \ldots, I_k$ and $k = 1, 2 \cdots$, are i.i.d. copies of $\{Z(t)\}_{t \geq 0}$ started with a single ancestor cell at time S_k.

When $\{U_i\}_{i=1}^{\infty}$ are i.i.d. exponentially distributed r.v. with c.d.f. $G_0(x) = P\{U_i \leq x\} = 1 - e^{-rx}$, $x \geq 0$, $\Pi(t)$ reduces to an ordinary Poisson process with cumulative rate $R(t) = rt$, and $Y(t)$ is an age-dependent branching process with homogeneous Poisson immigration. This process was investigated by Yanev [31].

Introduce the p.g.f. $\Psi(t; s) = E[s^{Y(t)} | Y(0) = 0]$. Yakovlev and Yanev ([34]; Theorem 1) proved that $\Psi(t; s)$ admits the expression

$$\Psi(t; s) = \exp\left\{ -\int_0^t r(t-u)[1 - g(F(u; s))]du \right\}, \tag{9.5}$$

where $F(u; s)$ satisfies the integral equation (9.1), and it satisfies the initial condition $\Psi(0, s) = 1$. Define the moments of the process with immigration

$$M(t) = E[Y(t)|Y(0) = 0] = \left. \frac{\partial \Psi(t; s)}{\partial s} \right|_{s=1},$$

$$M_2(t) = E[Y(t)(Y(t) - 1)|Y(0) = 0] = \left. \frac{\partial^2 \Psi(t; s)}{\partial s^2} \right|_{s=1},$$

$$W(t) = \mathrm{Var}[Y(t)|Y(0) = 0] = M_2(t) + M(t)(1 - M(t)).$$

We deduce from Eq. (9.5) that

$$M(t) = \gamma \int_0^t r(t-u)A(u)du, \tag{9.6}$$

$$\begin{aligned} M_2(t) = \gamma & \int_0^t r(t-u)B(u)du \\ & + [\gamma \int_0^t r(t-u)A(u)du]^2 + \gamma_2 \int_0^t r(t-u)A^2(u)du, \end{aligned}$$

$$W(t) = \int_0^t r(t-u)[\gamma V(u) + (\gamma + \gamma_2)A^2(u)]du. \tag{9.7}$$

To derive expressions for the covariances, define the joint p.g.f.

$$\Psi(s_1, s_2; t, \tau) = E[s_1^{Y(t)} s_2^{Y(t+\tau)} | Y(0) = 0], \quad t, \tau \geq 0.$$

Using the same line of arguments as Yakovlev and Yanev [34] to prove identity (9.5), we find that

$$\Psi(s_1,s_2;t,\tau)=\exp\left\{-\int_0^t r(u)[1-g(F(s_1,s_2;t-u,\tau))]du\right.$$
$$\left.-\int_t^{t+\tau} r(v)[1-g(F(1,s_2;t,\tau-v))]dv\right\}, \qquad (9.8)$$

where $F(s_1,s_2;t,\tau)=E[s_1^{Z(t)}s_2^{Z(t+\tau)}]$ is well determine by Eq. (22) in Sevastyanov [29], Chap. VIII.8, as follows

$$F(s_1,s_2;t,\tau)=\int_o^{t+0} h(u;F(s_1,s_2;t-u,\tau))dG(u)$$
$$+s_1\int_{t+0}^{t+\tau+0} h(u;F(s_2,t+\tau-u))dG(u)+s_1s_2(1-G(t+\tau)).$$

For the covariance

$$C(t,\tau)=\mathrm{Cov}[Y(t),Y(t+\tau)]=\left.\frac{\partial^2\log\Psi(s_1,s_2;t,\tau)}{\partial s_i\partial s_j}\right|_{s_1=s_2=1}$$

we deduce from Eq. (9.8) that

$$C(t,\tau)=\int_0^t r(u)[\gamma B(t-u,\tau)+\gamma_2 A(t-u)A(t+\tau-u)]du, \qquad (9.9)$$

with the initial condition $B(0,\tau)=B(\tau)$.

9.4 Asymptotic Expansions for the Moments

Lemma 9.1 *Let $a(t)$ and $b(t)$ be nonnegative functions defined on $[0,\infty)$. Assume that $a(t)$ is bounded, $\lim_{t\to\infty} a(t)=a<\infty$ and $\int_0^\infty b(t)dt=b<\infty$. Then*

$$\lim_{t\to\infty}\int_0^t a(u)b(t-u)du=ab.$$

Proof We have, for every $0<\Delta<1$, that

$$\int_0^t a(u)b(t-u)du=\int_0^{\Delta t} a(u)b(t-u)du+\int_{\Delta t}^t a(u)b(t-u)du=I_1(t)+I_2(t).$$

Note first that, for large enough t,

$$I_1(t) \leq \max_{0 \leq u \leq \Delta t} a(u) \int_{t(1-\Delta)}^{t} b(u)du \to 0, \quad t \to \infty.$$

and, for every $\varepsilon > 0$:

$$(a - \varepsilon) \int_0^{t(1-\Delta)} b(u)du \leq I_2(t) \leq (a + \varepsilon) \int_0^{t(1-\Delta)} b(u)du,$$

which completes the proof. □

Proposition 9.1 *Define* $\hat{r}_t(\alpha) = \int_0^t r(u)e^{-\alpha u}du$ *and assume that*

$$\lim_{t \to \infty} \hat{r}_t(\alpha) = \hat{r}(\alpha) < \infty. \tag{9.10}$$

Then, as $t \to \infty$, *we have that:*

$$M(t) = A\gamma\hat{r}(\alpha)e^{\alpha t}(1 + o(1)), \tag{9.11}$$

$$W(t) = e^{2\alpha t}\hat{r}(2\alpha)\left[\gamma V + (\gamma + \gamma_2)A^2\right](1 + o(1)), \tag{9.12}$$

$$C(t, \tau) = e^{\alpha\tau + 2\alpha t}\hat{r}(2\alpha)\left[\gamma B + \gamma_2 A^2\right](1 + o(1)). \tag{9.13}$$

Proof Manipulating Eq. (9.6) leads to

$$M(t) = \gamma e^{\alpha t} \int_0^t r(t-u)e^{-\alpha(t-u)}A(u)e^{-\alpha u}du.$$

Using Eqs. (9.2) and (9.10) and applying Lemma 9.1 we obtain as $t \to \infty$

$$\int_0^t r(t-u)e^{-\alpha(t-u)}A(u)e^{-\alpha u}du \to A\hat{r}(\alpha).$$

The above convergence obviously establishes (9.11).

To prove the asymptotic behavior of the variance $W(t)$, we use Eqs. (9.2), (9.4) and (9.7) to obtain

$$W(t) = e^{2\alpha t} \int_0^t \frac{r(t-u)}{e^{2\alpha(t-u)}} \left[\gamma \frac{V(u)}{e^{2\alpha u}} + (\gamma + \gamma_2)\frac{A^2(u)}{e^{2\alpha u}}\right] du.$$

Applying Lemma 9.1 again and using (9.10) leads to (9.12).

Finally we deduce from Eq. (9.9) that

$$C(t,\tau) = \int_0^t r(u)[\gamma B(t-u,\tau) + \gamma_2 A(t-u)A(t+\tau-u)]du$$
$$= e^{\alpha\tau+2\alpha t}\int_0^t \frac{r(u)}{e^{2\alpha u}}\left[\gamma \frac{B(t-u,\tau)}{e^{\alpha\tau+2\alpha(t-u)}} + \gamma_2 \frac{A(t-u)}{e^{\alpha(t-u)}}\frac{A(t+\tau-u)}{e^{\alpha(\tau+t-u)}}\right]du.$$

Now Eqs. (9.2), (9.3) and (9.10) and Lemma 9.1 lead to (9.13). □

Remark 9.1 Condition (9.10) is fulfilled if $r(t) = L(t)t^{\theta}$ for some s.v.f. $L(t)$ and some constant $\theta \in R$, or, more generally, if $r(t) = O(e^{\rho t})$ for some constant $\rho < \alpha$.

Proposition 9.2 *Assume that* $r(t) = re^{\rho t}$, $r > 0$ *and* $\rho > 0$. *Then, as* $t \to \infty$,

(i) $M(t) = \gamma r\hat{A}(\rho)e^{\rho t}(1+o(1))$, *if* $\rho > \alpha$, $\hat{A}(\rho) = \int_0^\infty e^{-(\rho-\alpha)u}A(u)e^{-\alpha u}du < \infty$;

(ii) $M(t) = \gamma rAte^{\alpha t}(1+o(1))$, *if* $\rho = \alpha$;

(iii) $M(t) = \dfrac{A\gamma r}{\alpha-\rho}e^{\alpha t}(1+o(1))$, *if* $\rho < \alpha$.

Proof In case (i), we deduce from Eq. (9.6) that

$$M(t) = \gamma re^{\rho t}\int_0^t e^{-(\rho-\alpha)u}A(u)e^{-\alpha u}du.$$

Since $A(u)e^{-\alpha u} \to A$ as $u \to \infty$ (see Eq. (9.2)), and $\rho - \alpha > 0$ the above integral converges to the constant $\hat{A}(\rho) = \int_0^\infty e^{-(\rho-\alpha)u}A(u)e^{-\alpha u}du$. This completes the proof.

To prove the result stated in case (ii), we write

$$M(t) = \gamma re^{\alpha t}\int_0^t A(u)e^{-\alpha u}du.$$

Since $A(u)e^{-\alpha u} \to A$ as $u \to \infty$, then $\int_0^t A(u)e^{-\alpha u}du \sim At$ as $t \to \infty$, which completes the proof.

Case (iii) follows from (1.11) with $\hat{r}(\alpha) = \frac{1}{\alpha-\rho}$. □

Proposition 9.3 *Assume that* $r(t) = re^{\rho t}$, $r > 0$ *and* $\rho > 0$. *Then, as* $t \to \infty$,

(i) $W(t) = We^{\rho t}(1+o(1))$ *if* $\rho > 2\alpha$,

$$W = r\int_0^\infty e^{-\rho u}[\gamma V(u) + (\gamma+\gamma_2)A^2(u)]du;$$

(ii) $W(t) = r[\gamma V + (\gamma+\gamma_2)A^2]te^{2\alpha t}(1+o(1))$ *if* $\rho = 2\alpha$;

(iii) $W(t) = \dfrac{r[\gamma V + (\gamma+\gamma_2)A^2]}{2\alpha-\rho}e^{2\alpha t}(1+o(1))$ *if* $\rho < 2\alpha$.

Proof In case (i), we use Eq. (9.7) to write

$$W(t) = re^{\rho t}\int_0^t e^{-(\rho-2\alpha)u}[\gamma V(u) + (\gamma+\gamma_2)A^2(u)]e^{-2\alpha u}du. \tag{9.14}$$

It follows from Eqs. (9.2) and (9.4) that

$$\left[\gamma V(u) + (\gamma+\gamma_2)A^2(u)\right]e^{2\alpha u} \to \left[\gamma V + (\gamma+\gamma_2)A^2\right], u \to \infty.$$

Since $\rho > 2\alpha$, the integral in Eq. (9.14) converges. This completes the proof of (i).

In case (ii), where $\rho = 2\alpha$, we have

$$\begin{aligned} W(t) &= re^{2\alpha t}\int_0^t e^{-2\alpha u}[\gamma V(u) + (\gamma+\gamma_2)A^2(u)]du \\ &\sim r[\gamma V + (\gamma+\gamma_2)A^2]te^{2\alpha t}(1+o(1)) \end{aligned}$$

as $t \to \infty$, and the result follows.

In case (iii), where $\rho < 2\alpha$, we have

$$\begin{aligned} W(t) &= \int_0^t re^{\rho u}\frac{[\gamma V(t-u) + (\gamma+\gamma_2)A^2(t-u)]}{[\gamma V + (\gamma+\gamma_2)A^2]e^{2\alpha(t-u)}}[\gamma V + (\gamma+\gamma_2)A^2]e^{2\alpha(t-u)}du \\ &= r[\gamma V + (\gamma+\gamma_2)A^2]e^{2\alpha t}\int_0^t e^{(\rho-2\alpha)u}\frac{[\gamma V(t-u) + (\gamma+\gamma_2)A^2(t-u)]}{[\gamma V + (\gamma+\gamma_2)A^2]e^{2\alpha(t-u)}}du. \end{aligned}$$

Since $\int_0^\infty e^{(\rho-2\alpha)t}dt = \dfrac{1}{2\alpha-\rho}$ and $\dfrac{[\gamma V(t) + (\gamma+\gamma_2)A^2(t)]}{[\gamma V + (\gamma+\gamma_2)A^2]e^{2\alpha(t)}} \to 1$ as $t \to \infty$, the result follows by Lemma 9.1. □

Proposition 9.4 *Assume that* $r(t) = re^{\rho t}$, $r > 0$ *and* $\rho > 0$. *Then, as* $t \to \infty$,

(i) $C(t,\tau) = Cre^{\alpha\tau+2\rho t}(1+o(1))$ *if* $\rho > 2\alpha$,

$$C = \int_0^\infty e^{-(\rho-2\alpha)u}\left[\gamma\frac{B(u,\tau)}{e^{\alpha\tau+2\alpha u}} + \gamma_2\frac{A(u)}{e^{\alpha u}}\frac{A(\tau+u)}{e^{\alpha(\tau+u)}}\right]du,$$

.

(ii) $C(t,\tau) = r[\gamma B + \gamma_2 A^2]te^{\alpha\tau+2\alpha t}(1+o(1))$ *if* $\rho = 2\alpha$.

(iii) $C(t,\tau) = \dfrac{r[\gamma B + \gamma_2 A^2]}{2\alpha-\rho}e^{\alpha\tau+2\alpha t}(1+o(1))$ *if* $\rho < 2\alpha$.

Proof We obtain from Eq. (9.9) that

$$C(t,\tau) = re^{\alpha\tau+2\rho t}\int_0^t e^{-(\rho-2\alpha)u}\left[\gamma\frac{B(u,\tau)}{e^{\alpha\tau+2\alpha u}} + \gamma_2\frac{A(u)}{e^{\alpha u}}\frac{A(\tau+u)}{e^{\alpha(\tau+u)}}\right]du.$$

In case (i) as $u \to \infty$, we deduce from (9.2) and (9.3) that

$$\left[\gamma \frac{B(u,\tau)}{e^{\alpha\tau+2\alpha u}} + \gamma_2 \frac{A(u)}{e^{\alpha u}} \frac{A(\tau+u)}{e^{\alpha(\tau+u)}}\right] \to \gamma B + \gamma_2 A^2 \tag{9.15}$$

Since $\rho > 2\alpha$, C is finite which completes the proof.

In case (ii), we deduce from Eq. (9.15) that

$$\int_0^t \left[\gamma \frac{B(u,\tau)}{e^{\alpha\tau+2\alpha u}} + \gamma_2 \frac{A(u)}{e^{\alpha u}} \frac{A(\tau+u)}{e^{\alpha(\tau+u)}}\right] du \sim [\gamma B + \gamma_2 A^2]t,$$

as $t \to \infty$, which completes the proof of (ii).

In case (iii), we deduce from Eq. (9.9) that

$$\begin{aligned} C(t,\tau) &= \int_0^t re^{\rho u} \left[\gamma \frac{B(t-u,\tau)}{e^{\alpha\tau+2\alpha(t-u)}} + \gamma_2 \frac{A(t-u)}{e^{\alpha(t-u)}} \frac{A(\tau+t-u)}{e^{\alpha(\tau+t-u)}}\right] e^{\alpha\tau+2\alpha(t-u)} du \\ &= re^{\alpha\tau+2\alpha t} \int_0^t e^{(\rho-2\alpha)u} \left[\gamma \frac{B(t-u,\tau)}{e^{\alpha\tau+2\alpha(t-u)}} + \gamma_2 \frac{A(t-u)}{e^{\alpha(t-u)}} \frac{A(\tau+t-u)}{e^{\alpha(\tau+t-u)}}\right] du. \end{aligned}$$

Since $\int_0^\infty e^{(\rho-2\alpha)t} dt = \frac{1}{2\alpha-\rho}$ and $\left[\gamma \frac{B(t,\tau)}{e^{\alpha\tau+2\alpha t}} + \gamma_2 \frac{A(t)}{e^{\alpha t}} \frac{A(\tau+t)}{e^{\alpha(\tau+t)}}\right] \to [\gamma B + \gamma_2 A^2]$, as $t \to \infty$ then the result follows by Lemma 9.1. □

9.5 Limit Theorems

Theorem 9.1 *Assume that* $\hat{r}(\alpha) = \lim_{t\to\infty} \hat{r}_t(\alpha) < \infty$. *Then, as* $t \to \infty$,

$$\zeta(t) = Y(t)/M(t) \xrightarrow{L_2} \zeta,$$

where ζ *is a random variable with expectation* $E\zeta = 1$ *and variance*

$$Var(\zeta) = \hat{r}(2\alpha)\left[\gamma V + (\gamma+\gamma_2)A^2\right][A\gamma\hat{r}(\alpha)]^{-2}.$$

Proof Note first that $E[\zeta(t)] \equiv 1$. Hence, it is sufficient to prove that

$$\Delta(t,\tau) = E[\zeta(t+\tau) - \zeta(t)]^2 \to 0 \text{ as } t \to \infty,$$

uniformly for $\tau \geq 0$. We have

$$\Delta(t,\tau) = \mathrm{Var}[\zeta(t+\tau)] + \mathrm{Var}[\zeta(t)] - 2\mathrm{Cov}[\zeta(t), \zeta(t+\tau)] \tag{9.16}$$

where

$$\mathrm{Var}[\zeta(t)] = \frac{W(t)}{M^2(t)} \quad \text{and} \quad \mathrm{Cov}[\zeta(t), \zeta(t+\tau)] = \frac{C(t,\tau)}{M(t)M(t+\tau)}. \tag{9.17}$$

As $t \to \infty$, we deduce from Eqs. (9.11)–(9.13) that

$$\begin{aligned}\mathrm{Var}[\zeta(t)] &= \frac{e^{2\alpha t}\hat{r}(2\alpha)\left[\gamma V + (\gamma+\gamma_2)A^2\right]}{[A\gamma\hat{r}(\alpha)]^2\, e^{2\alpha t}}(1+o(1)) \\ &= \hat{r}(2\alpha)\left[\gamma V + (\gamma+\gamma_2)A^2\right][A\gamma\hat{r}(\alpha)]^{-2}(1+o(1)),\end{aligned}$$

$$\begin{aligned}\mathrm{Var}[\zeta(t+\tau)] &= \frac{e^{2\alpha(t+\tau)}\hat{r}(2\alpha)\left[\gamma V + (\gamma+\gamma_2)A^2\right]}{[A\gamma\hat{r}(\alpha)]^2\, e^{2\alpha(t+\tau)}}(1+o(1)) \\ &= \hat{r}(2\alpha)\left[\gamma V + (\gamma+\gamma_2)A^2\right][A\gamma\hat{r}(\alpha)]^{-2}(1+o(1)),\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}[\zeta(t), \zeta(t+\tau)] &= \frac{e^{\alpha\tau+2\alpha t}\hat{r}(2\alpha)\left[\gamma B + \gamma_2 A^2\right]}{[A\gamma\hat{r}(\alpha)]^2\, e^{2\alpha t+\alpha\tau}}(1+o(1)) \\ &= \hat{r}(2\alpha)\left[\gamma B + \gamma_2 A^2\right][A\gamma\hat{r}(\alpha)]^{-2}(1+o(1)).\end{aligned}$$

The result stated in the theorem follows using the fact that $V + A^2 - B = 0$. ☐

Theorem 9.2 *Assume that $r(t) = re^{\rho t}$ with $\rho \geq \alpha$ and $r > 0$. Then, as $t \to \infty$,*

$$\zeta(t) = Y(t)/M(t) \xrightarrow{L_2} 1 \ \ \textit{and} \ \ \zeta(t) \xrightarrow{a.s.} 1.$$

Proof Assume first that $\rho = \alpha$. Then $\rho < 2\alpha$, and we deduce from Propositions 9.2.(ii), 9.3.(iii), and 9.4.(iii), as $t \to \infty$, that

$$M(t) = \gamma r A t e^{\alpha t}(1+o(1)), \tag{9.18}$$

$$W(t) = \frac{r}{2\alpha-\rho}\left[\gamma V + (\gamma+\gamma_2)A^2\right]e^{2\alpha t}(1+o(1)), \tag{9.19}$$

$$C(t,\tau) = r\left[\gamma B + \gamma_2 A^2\right]e^{\alpha\tau+2\alpha t}(1+o(1)). \tag{9.20}$$

Note that $E[\zeta(t)] \equiv 1$. We deduce from Eqs. (9.18)–(9.20), as $t \to \infty$, that

$$\mathrm{Var}[\zeta(t)] = \frac{r[\gamma V + (\gamma+\gamma_2)A^2]e^{2\alpha t}}{(2\alpha-\rho)(\gamma r A)^2 t^2 e^{2\alpha t}}(1+o(1)) = O(t^{-2}),$$

$$\mathrm{Var}[\zeta(t+\tau)] = \frac{r[\gamma V + (\gamma+\gamma_2)A^2]e^{2\alpha(t+\tau)}}{(2\alpha-\rho)(\gamma r A)^2 (t+\tau)^2 e^{2\alpha(t+\tau)}}(1+o(1)) = O(t^{-2}),$$

$$\mathrm{Cov}[\zeta(t), \zeta(t+\tau)] = \frac{r[\gamma B + \gamma_2 A^2]e^{\alpha\tau+2\alpha t}}{(2\alpha-\rho)(\gamma r A)^2 t(t+\tau)e^{\alpha\tau+2\alpha t}}(1+o(1)) = O(t^{-2}).$$

Now the first relation follows from $E[\zeta(t)] = 1$ and $\mathrm{Var}[\zeta(t)] \to 0$ as $t \to \infty$. Next, Eqs. (9.16), (9.17) and the above three identities imply that

$$\Delta(t,\tau) := E[\zeta(t+\tau) - \zeta(t)]^2 = O(t^{-2}),$$

uniformly on $\tau \geq 0$, as $t \to \infty$.

Hence, $\int_0^\infty \Delta(t,\tau)dt < \infty$, and Theorem 21.1 in Harris [7] entails that $\zeta(t)$ converges almost surely to 1.

The proofs of the remaining cases are similar. The result follows from the asymptotic behavior of $\Delta(t,\tau)$:

$$\Delta(t,\tau) = \begin{cases} O(t^{-2}) & \text{if } \rho = \alpha \\ O(e^{-2(\alpha-\rho)t}) & \text{if } \alpha < \rho < 2\alpha \\ O(e^{-2\alpha t}) & \text{if } \rho = 2\alpha \\ O(e^{-\rho t}) & \text{if } \rho > 2\alpha. \end{cases}$$

□

Remark 9.2 Theorem 9.2 can be interpreted as a SLLN. Hence one can conjecture a CLT.

Theorem 9.3 *Assume that $r(t) = re^{\rho t}$ with $\rho \geq \alpha$ and $r > 0$.*

(i) If $\alpha \leq \rho \leq 2\alpha$, then $X(t) = \dfrac{Y(t) - M(t)}{\sqrt{W(t)}} \xrightarrow{d} N(0,1)$ as $t \to \infty$.

(ii) If $\rho > 2\alpha$, then $X(t) \xrightarrow{d} N(0,\sigma^2)$ as $t \to \infty$,
where $\sigma^2 = 1 - \gamma r\hat{A}(\rho)\left\{r\int_0^\infty e^{-\rho u}[\gamma V(u) + (\gamma+\gamma_2)A^2(u)]du\right\}^{-1}$.

Proof Define the characteristic function $\varphi_t(z) := E[e^{izX(t)}]$. It follows from the definition of $X(t)$ that

$$\varphi_t(z) = e^{-izM(t)/\sqrt{W(t)}}\Psi(t; e^{iz/\sqrt{W(t)}}),$$

where $\Psi(t;s) := E[s^{Y(t)}|Y(0) = 0]$. Next, we deduce from Eq. (9.5) that

$$\log\varphi_t(z) = -izM(t)/\sqrt{W(t)} - \int_0^t r(t-u)[1 - g(F(u; e^{iz/\sqrt{W(t)}}))]du.$$

Note that, as $s \to 1$,

$$1 - g(s) \sim \gamma(1-s) - \gamma_2(1-s)^2/2,$$
$$1 - F(u;s) \sim A(u)(1-s) - B(u)(1-s)^2/2.$$

and, as $x \to 0$, $1 - e^{cx} \sim -cx$. Therefore, as $t \to \infty$,

$$\log \varphi_t(z) \sim -izM(t)/\sqrt{W(t)} \\ - \int_0^t r(t-u)\{\gamma[1 - F(u; e^{iz/\sqrt{W(t)}})] - \gamma_2[1 - F(u; e^{iz/\sqrt{W(t)}}]^2/2\}du. \tag{9.21}$$

As $t \to \infty$,

$$1 - F(u; e^{iz/\sqrt{W(t)}}) \sim A(u)(1 - e^{iz/\sqrt{W(t)}}) - B(u)(1 - e^{iz/\sqrt{W(t)}})^2/2 \\ \sim -izA(u)/\sqrt{W(t)} + z^2 B(u)/\sqrt{W(t)}/2.$$

Hence,

$$D(t) = \int_0^t r(t-u)\left\{\gamma[1 - F(u; e^{iz/\sqrt{W(t)}})] - \gamma_2[1 - F(u; e^{iz/\sqrt{W(t)}}]^2/2\right\} du \\ \sim -iz\gamma \int_0^t r(t-u)A(u)du/\sqrt{W(t)} + (z^2/2)\gamma \int_0^t r(t-u)B(u)du/W(t) \\ + (z^2/2)\gamma_2 \int_0^t r(t-u)A^2(u)du/W(t).$$

Now using (9.6) and (9.7) one has as $t \to \infty$,

$$D(t) \sim -izM(t)/\sqrt{W(t)} + (z^2/2)[1 - M(t)/W(t)].$$

Hence one gets from (9.21)

$$\log \varphi_t(z) \sim -(z^2/2)[1 - M(t)/W(t)], \; t \to \infty. \tag{9.22}$$

As $t \to \infty$, we deduce from the asymptotic of $M(t)$ and $W(t)$ (see Sect. 9.4) that $M(t)/W(t) \to 0$ when $\alpha \le \rho \le 2\alpha$ and $M(t)/W(t) \to \frac{\gamma r\hat{A}(\rho)}{W}$ when $\rho > 2\alpha$. Then $\sigma^2 = 1 - \frac{\gamma r\hat{A}(\rho)}{W}$.

Therefore, returning to Eq. (9.22), we finally obtain that $\lim_{t\to\infty} \varphi_t(z) = e^{-z^2/2}$ for $\alpha \le \rho < 2\alpha$ and $\lim_{t\to\infty} \varphi_t(z) = e^{-z^2\sigma^2/2}$ for $\rho > 2\alpha$ which are characteristic functions of normal distributions. Assertions (*i*) and (*ii*) follow by applying the continuity theorem [5]. ☐

Remark 9.3 Recall that the relation

$$X(t) = [Y(t) - M(t)]/\sqrt{W(t)} \xrightarrow{d} N(0, \sigma^2) \text{ as } t \to \infty$$

is often presented as $Y(t) \in N(M(t), \sigma^2 W(t))$ and one can say that $Y(t)$ has asymptotic normality with a mean $M(t)$ and a variance $\sigma^2 W(t)$. Then from Theorem 9.3

using Propositions 9.2 and 9.3 one can obtain the following relations which give more convenient interpretation for the rate of convergence:

(a) If $\rho = \alpha$ then $Y(t)/te^{\alpha t} \in N\left(\gamma r A, \frac{K}{(2\alpha-\rho)t^2}\right)$,

(b) If $\alpha < \rho < 2\alpha$ then $Y(t)/e^{\rho t} \in N\left(\gamma r \hat{A}(\rho), \frac{K}{(2\alpha-\rho)} e^{2(\alpha-\rho)t}\right)$,

(c) If $\rho = 2\alpha$ then $Y(t)/e^{2\alpha t} \in N\left(\gamma r \hat{A}(2\alpha), Kte^{-2\alpha t}\right)$,

(d) If $\rho > 2\alpha$ then $Y(t)/e^{\rho t} \in N(\gamma r \hat{A}(\rho), \sigma^2 W e^{-\rho t})$,

where $K = r(\gamma + (\gamma + \gamma_2)A^2)$ and $W = r \int_0^\infty e^{-\rho u}[\gamma V(u) + (\gamma + \gamma_2)A^2(u)]du$.

Note that these relations are also useful for constructing of asymptotically confident intervals.

Acknowledgements The research was partially supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No. DFNI I02/17. The authors would like to thank the referee for the comments.

References

1. Ahsanullah, M., Yanev, G.P.: Records and Branching Processes. Nova Science publishers, New York (2008)
2. Asmussen, S., Hering, H.: Branching Processes. Birkhauser, Boston (1983)
3. Athreya, K.B., Ney, P.E.: Branching Processes. Springer, New York (1972)
4. Chen, R., Hyrien, O., Mayer-Proschel M., Noble, M.: A composite likelihood approach to the analysis of longitudinal clonal data under an age-dependent branching process. Biostatistics **12**, 173–191 (2010)
5. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 2. Wiley, New York (1971)
6. Haccou, P., Jagers, P., Vatutin, V.A.: Branching Processes: Variation, Growth and Extinction of Populations. Cambridge University Press, Cambridge (2005)
7. Harris, T.E.: The Theory of Branching Processes. Springer, Berlin (1963)
8. Hyrien, O., Yanev, N.M.: Modeling cell kinetics using branching processes with non-homogeneous Poisson immigration. C. R. Acad. Bulgare Sci. **63**, 1405–1414 (2010)
9. Hyrien, O., Yanev, N.M.: Age-dependent branching processes with non-homogeneous Poisson immigration as models of cell kinetics. In: Oakes, D., Hall, W.J., Almudevar, A. (eds.) Modeling and Inference in Biomedical Sciences: In Memory of Andrei Yakovlev. IMS Collections Series. Institute of Mathematical Statistics, Beachwood, OH (2012, in press)
10. Hyrien, O., Mayer-Proschel, M., Noble, M., Yakovlev, A.: A stochastic model to analyze clonal data on multi type cell populations. Biometrics **61**, 199–207 (2005)
11. Hyrien, O., Chen, R., Mayer-Proschel, M., Noble, M.: Saddlepoint approximations to the moments of age-dependent branching processes, with applications. Biometrics **66**, 567–577 (2010)
12. Hyrien, O., Dietrich, J., Noble, M.: Quantitative analysis of the effect of chemotherapy on division and differentiation of glial precursor cells. Cancer Res. **70**(24), 10051–10059 (2010)
13. Hyrien, O., Peslak, S., Yanev, N.M., Palis, J.: A stochastic model of stress erythropoiesis using a two-type age-dependent branching process with immigration. J. Math. Biol. **70**, 1485–1521 (2015)

14. Jagers, P.: Age-dependent branching processes allowing immigration. Theory Probab. Appl. **13**, 225–236 (1968)
15. Jagers, P.: Branching Processes with Biological Applications. Wiley, London (1975)
16. Kaplan, N., Pakes, A.G.: Supercritical age-dependent branching processes with immigration. Stoch. Process. Appl. **2**, 371–389 (1974)
17. Kimmel, M., Axelrod, D.E.: Branching Processes in Biology. Springer, New York (2002)
18. Mitov, K.V., Yanev, N.M.: Bellman–Harris branching processes with state-dependent immigration. J. Appl. Probab. **22**, 757–765 (1982)
19. Mitov, K.V., Yanev, N.M.: Bellman–Harris branching processes with a special type of state-dependent immigration. Adv. Appl. Probab. **21**, 270–283 (1989)
20. Mitov, K.V., Yanev, N.M.: Critical Bellman–Harris branching processes with infinite variance allowing state-dependent immigration. Stoch. Models **18**, 281–300 (2002)
21. Mitov, K.V., Yanev, N.M.: Sevastyanov branching processes with non-homogeneous poisson immigration. Proc. Steklov Inst. Math. **282**, 172–185 (2013)
22. Mode, C.J.: Branching Processes. Springer, New York (1972)
23. Pakes, A.G.: Limit theorems for an age-dependent branching process with immigration. Math. Biosci. **14**, 221–234 (1972)
24. Pakes, A.G., Kaplan, N.: On the subcritical Bellman–Harris process with immigration. J. Appl. Probab. **11**, 652–668 (1974)
25. Radcliffe, J.: The convergence of a super-critical age-dependent branching processes allowing immigration at the epochs of a renewal process. Math. Biosci. **14**, 37–44 (1972)
26. Sevastyanov, B.A.: Limit theorems for branching random processes of special type. Theory Probab. Appl. **2**, 339–348 (1957) (in Russian)
27. Sevastyanov, B.A.: Age-dependent branching processes. Theory Probab. Appl. **9**, 577–594 (1964). (Letter **11**, 363–364 (1966)) (in Russian)
28. Sevastyanov, B.A.: Limit theorems for age-dependent branching processes. Theory Probab. Appl. **13**, 243–265 (1968) (in Russian)
29. Sevastyanov, B.A.: Branching Processes. Nauka, Moscow (1971) (in Russian)
30. Yanev, N.M.: Branching stochastic processes with immigration. Bull. Inst. Math. Acad. Bulgare Sci. **15**, 71–88 (1972) (in Bulgarian)
31. Yanev, N.M.: On a class of decomposable age-dependent branching processes. Math. Balk. **2**, 58–75 (1972) (in Russian)
32. Yakovlev, A.Y., Yanev, N.M.: Transient Processes in Cell Proliferation Kinetics. Springer, Berlin (1989)
33. Yakovlev, A.Y., Yanev, N.M.: Branching stochastic processes with immigration in analysis of renewing cell populations. Math. Biosci. **203**, 37–63 (2006)
34. Yakovlev, A.Y., Yanev, N.M.: Age and residual lifetime distributions for branching processes. Stat. Probab. Lett. **77**, 503–513 (2007)

Chapter 10
Crump-Mode-Jagers Branching Process: A Numerical Approach

Plamen Trayanov

Mathematics Subject Classification (2000): 60J80

10.1 Introduction

The Crump-Mode-Jagers branching process or also called General Branching Process (GBP) describes a population in which each woman could have random life length and random intervals of time between each birth (see Jagers [5]). All individuals are assumed to have the same birth and death distributions regardless of the time. The expected population count for such branching process satisfies a specific renewal equation. Its theoretical solution is given by renewal theory. As explained in Mitov and Omey [8] the solution is a convolution of two functions, one of which is the renewal function, which is an infinite sum of convolutions with increasing order. Calculating even one of the terms in this sum numerically is computationally heavy and very time consuming. To make things worse, in order to find the expected future age structure we need to solve a 100 renewal equations—one for each age. It is preferable to use a more suitable numerical method for solving the renewal equation in order to reduce computation time. Some examples of such numerical methods can be found in [1, 4, 9, 13].

This paper presents a numerical approach for projecting the population age structure based on the theory of General Branching Processes that consists only of simple matrix multiplications and solves all the required renewal equations at the

---

P. Trayanov (✉)
Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", 5, J. Bourchier Blvd, 1164 Sofia, Bulgaria
e-mail: plamentrayanov@gmail.com

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_10

same time. A special case of this numerical method turns out to be the Leslie matrix projection, widely used in demographics. The estimation error of the expected total population count, calculated using Leslie matrix is proved to be of order $O(h)$ and the estimation error for each particular age interval—of order $O(h^2)$. The result justifies the use of Leslie matrix in demographics. From another point of view the Leslie matrix projection turns out to be a very effective numerical method for solving renewal equations.

Section 10.2 reviews the definition of GBP and some of the results presented in Jagers [5]. Section 10.3 reviews a GBP that is relevant for human population and some theoretical results are derived. The Appendix reviews two examples. The first example shows the application of the numerical method for projecting the population of Bulgaria given the current age structure. The second example shows the application of this numerical method for solving renewal equations.

10.2 Preliminary Results

For consistency, the notation used in this paper is the same as in Jagers [5].

Let I be the set of all n-tuples of non-negative integers for all n. This is an index set for individuals in the population. The GBP presented in Jagers [5] assumes the process starts from a single individual denoted by (0). The woman (0) is assumed to have age 0 at time $t = 0$. The n-th child of $x \in I$ is denoted by (x, n).

We denote by ξ a point process defined on R^+ and $\mu(A) = \mathbb{E}(\xi(A))$, where A is a Borel set in R^+. The individual $(x, n) \in I$ exists if $\xi_x(\infty) \geq n$. The point process models the births of a woman.

Let λ_x be a random variable that models the life length of the individual x and ξ_x be a point process of a woman x. For each $x \in I$ is defined a couple (λ_x, ξ_x) and these couples are assumed to be independent and identically distributed. This means the distributions of ξ and λ are time invariant.

Let $\tau_x(k) = \inf\{t : \xi_x(t) \geq k\}$ be the age of birth of child (x, k). Let $\sigma_x = \tau_0(j_1) + \tau_{j_1}(j_2) + \ldots + \tau_{(j_1,\ldots,j_{n-1})}(j_n)$, where $x = (j_1, \ldots, j_n)$ and $\sigma_0 = 0$.

Definition 10.1 The GBP is defined as $z_t^a = \sum_{x \in I} z_t^a(x)$, where $z_t^a(x)$ is an indicator variable for the event that the individual x is alive and younger than $a > 0$ (or at age a) at time $t > 0$:

$$z_t^a(x) = \begin{cases} 1, \text{ when } t - a \leq \sigma_x \leq t < \sigma_x + \lambda_x, \\ 0, \text{ otherwise.} \end{cases}$$

At time t, the oldest individual is younger than t, so we can write $z_t = z_t^a$, when $a > t$. Note that the above definition includes individuals which age is exactly a in the branching process z_t^a, as opposed to the definition presented in Jagers [5].

Let $\mu(t) = \mu([0,t])$ for $t > 0$, $F(s) = \mathbb{E}(s^{\xi(\infty)})$, $|s| \leq 1$, $L(t) = \mathbb{P}(\lambda_x \leq t)$ and $S(t) = 1 - L(t)$. This latter denotes the survival probability function. It is known (see Jagers [5]) that if $F(s) < \infty$, $|s| \leq 1$, then $m_t = \mathbb{E}(z_t) < \infty$, for all t, moreover $m_t^a = \mathbb{E}(z_t^a)$ satisfies

$$m_t^a = 1_{[0,a]}(t)\{1 - L(t)\} + \int_0^t m_{t-u}^a \,\mu(\mathrm{d}u). \tag{10.1}$$

When $a = \infty$ it is skipped in the notation.

The theoretical solution of Eq. (10.1) is provided by renewal theory (see Mitov and Omey [8]). If we have an analytical expression for the functions S and μ then sometimes it is possible to find the theoretical solution of this equation but, in general, this is not the case. In the case of human populations the functions S and μ could be estimated from empirical data using smoothing splines and thus not have a decent analytical expression. Moreover, in practice, the theoretical functions that were researched as models of the mortality distribution in practice do not fit the data well enough and have some fitting problems due to the large number of parameters (a very impressive theoretical model of mortality can be found in Mode [7]). But even if we assume some analytical expression for the functions $S(t)$ and $\mu(t)$, solving the renewal equation could be very difficult. Another problem arises from the need to forecast not only the total population but the age structure too. If we have 100 age intervals (0-1, 1-2, ..., 99-100), then in order to calculate the age structure we need to solve 100 equations, substituting $a = 1, 2, 3, \ldots, 100$ in Eq. (10.1) and then taking the differences $m_t^{a+1} - m_t^a$.

10.3 Results

Let ${}_bz_t$ be a branching process started with a woman aged b at time $t = 0$, ${}_b\xi$ be the point process, ${}_b\mu$ be the expectation of the point process and ${}_bS$ be the survivability function. We have that ${}_b\mu(t) = \mathbb{E}(\xi(t + b) - \xi(b) \mid \lambda > b), t > 0$ is the expected number of children that a woman has after age b if she has survived to that age and ${}_bS(t) = \mathbb{P}(\lambda > b + t \mid \lambda > b)$ is the probability she lives t years after age b if she has survived to age b. Let ${}_bz_t^a$ be the number of individuals younger than a at time t, that corresponds to a woman at aged b for $t = 0$. If $a = \infty$ and $b = 0$ we will skip a and b in the notation.

Theorem 10.1 *Let $1_{[0,a-b]}(t)$ be an indicator function such that $1_{[0,a-b]}(t) = 1$, if $t \in [0, a - b]$, with $a - b \geq 0$, and $1_{[0,a-b]}(t) = 0$ if $a - b < 0$ or $t > a - b$.*

If $F(s) < \infty, |s| \leq 1$, *and* $S(b) > 0$, *then* ${}_bm_t = \mathbb{E}({}_bz_t) < \infty$, *for all t, and the expected number of individuals who are younger than a at time t (or at age a), started with a woman at age b at time zero,* ${}_bm_t^a = \mathbb{E}({}_bz_t^a)$, *satisfies*

$$ {}_bm_t^a = {}_bS(t) 1_{[0,a-b]}(t) + \int_0^t m_{t-u}^a \, {}_b\mu(\mathrm{d}u), \tag{10.2} $$

where ${}_bS(t) = \frac{S(b+t)}{S(b)}$ *denotes the probability that a woman at age b to survive* $b+t$, ${}_b\mu(t) = \frac{\mu(t+b)-\mu(b)}{S(b)}$ *is the expectation of that point process and* m_{t-u}^a *is the solution of the previous Eq. (10.1).*

In addition if $S(t)$ *and* $\mu(t)$ *are twice differentiable on t, then* ${}_bm_t^a$ *is twice differentiable both on t for* $t \neq a-b$ *and on b for* $b \neq a$. *If* $a = \infty$ *then* ${}_bm_t$ *is twice differentiable for all* $t > 0$.

Proof The proof is quite similar to the proof of Eq. (10.1) presented in Jagers [5] but note that the life-length of the first woman (0) and the point process have different distribution from the ones of her children. The first woman at age b at time 0 has $\xi_0(t)$ children in the interval $[0, t]$. Let $z_{t-\tau(n)}^{[n]a}$ denotes the branching process of the n-th child of this woman (0). Then we have that ${}_bz_t^a = {}_bz_t^a(0) + \sum_{n=1}^{\xi_0(t)} z_{t-\tau(n)}^{[n]a}$. This sum is finite and if we use the linearity of expectation we obtain

$$ \begin{aligned} {}_bm_t^a = \mathbb{E}{}_bz_t^a &= \mathbb{P}({}_bz_t^a(0) = 1) + \mathbb{E}\left(\sum_{n=1}^{\xi_0(t)} z_{t-\tau(n)}^{[n]a}\right) \\ &= \mathbb{P}({}_bz_t^a(0) = 1) + \mathbb{E}\left(\int_0^t z_{t-u}^a \, \xi_0(\mathrm{d}u)\right) \\ &= {}_bS(t) 1_{[0,a-b]}(t) + \int_0^t m_{t-u}^a \, {}_b\mu(\mathrm{d}u), \end{aligned} $$

where ${}_b\mu(u) = \mathbb{E}\xi_0(u) = \mathbb{E}(\xi(u+b) - \xi(b) \mid \lambda > b)$.

If $S(t)$ is twice differentiable and b is such that $S(b) > 0$ then ${}_bS(t) = \frac{S(t+b)}{S(b)}$ is correctly defined and twice differentiable both on t and on b. We have that ${}_b\mu(t) = \frac{\mu(t+b)-\mu(b)}{S(b)}$, consequently, if $\mu(t)$ is twice differentiable then ${}_b\mu(t)$ is twice differentiable both on t and on b. We have that the convolution of two twice differentiable functions is also twice differentiable, hence, m_t^a as a solution to the renewal equation (10.1), is also twice differentiable on t when $t \neq a$. Due to the term $1_{[0,a-b]}(t)$ in Eq. (10.2), we have that ${}_bm_t^a$ is twice differentiable on t for $t > 0, t \neq a, t \neq a-b$. If we set $a = \infty$, then it is twice differentiable for all $t > 0$. We also have that due to the term $1_{[0,a-b]}(t)$ in Eq. (10.2), ${}_bm_t^a$ is twice differentiable on b, when $b \neq a$. □

Let ω be the maximum age of the life table, i.e. $\mathbb{P}(\lambda > \omega) = 0$. In other words, it is the maximum age that a person can live.

In order to introduce the age structure, it is needed to separate the population by age groups using intervals with length h. We can assume the maximum age ω is multiple of h. Let $N_t(bh;(b+1)h]$ be the number of women in the age interval $(bh;(b+1)h]$ at time t, i.e. the population age structure. For example, if the branching process started with one woman of age 0 at time 0, then we have $z_t = N_t[0;h] + \sum_{b=1}^{\omega/h-1} N_t(bh;(b+1)h]$ and $N_t(bh;(b+1)h] = z_t^{bh+h} - z_t^{bh}$.

The following theorem shows that the expected age structure at time $t+h$ can be calculated from the expected age structure at time t. The theorem holds for branching processes starting with one or finite number of individuals at different ages (including random ages at time 0).

Theorem 10.2 *Let the conditions of Theorem 10.1 hold. Let $S(t)$ and $\mu(t)$ be twice differentiable on $\mathbb{R}$ and $\mu(t) = 0$ for $t \le 0$. Let ${}_u\mu''(t)$ be a bounded function of t and u and $S''(t)$ a bounded function of t. Then the following numerical approximation holds for $h \to 0$:*

$$\mathbb{E}\left(N_{t+h}((b+1)h;(b+2)h]\right) = \mathbb{E}\left(N_t(bh;(b+1)h]\right) \cdot \left[{}_{bh}S(h) + O(h^2)\right],$$

$$\mathbb{E}\left(N_{t+h}[0;h)\right) = \sum_{b=0}^{\omega/h-1} \mathbb{E}\left(N_t(bh;(b+1)h]\right) \cdot {}_{bh}\mu(h) + \mathbb{E}\left(N_t[0;\omega]\right) \cdot O(h^2),$$

where b is an integer, $b \ge 0$.

As alternative

$$\mathbb{E}\left(N_{t+h}[0;h)\right) = \sum_{b=0}^{\omega/h-1} \left[\mathbb{E}\left(N_t(bh;(b+1)h]\right) \cdot {}_{bh}\mu'(0)h\right] + \mathbb{E}\left(N_t[0;\omega]\right) \cdot O(h^2).$$

The constant in the error terms $O(h^2)$ does not depend on the choice of b.

Proof Let $u > 0$. By substituting $t = h$ in Theorem 10.1 and using the rectangular approximation of integrals, we can calculate the expected number of newborns and the expected number of individuals at age u that will survive to age $u + h$:

$$\begin{aligned} {}_u m_h^h &= {}_u S(h) 1_{[0,h-u]}(h) + \int_0^h m_{h-v}^h \, {}_u\mu(\mathrm{d}v) = \int_0^h m_{h-v} \, {}_u\mu(\mathrm{d}v) \\ &= m_0 \cdot ({}_u\mu(h) - {}_u\mu(0)) + O(h^2) = {}_u\mu(h) + O(h^2), \end{aligned}$$

$$ {}_u m_h^{u+h} - {}_u m_h^h = {}_u S(h) 1_{[0,h]}(h) + \int_0^h m_{h-v}^{u+h} \, {}_u\mu(\mathrm{d}v) - \int_0^h m_{h-v}^h \, {}_u\mu(\mathrm{d}v) = {}_u S(h). \tag{10.3}$$

Note that ${}_u m_h^b = {}_u m_h^h$ for every $h \leq b < u + h$ so the number of individuals in the age interval $(h, u+h)$ is 0, i.e. after time h we can only have individuals at age $u+h$ or age $[0, h]$.

From Taylor's theorem, it follows that the error term $O(h^2)$ in Eq. (10.3) has the form $\left[-m'_{h-\zeta} \cdot {}_u\mu'(\zeta) + m_{h-\zeta} \cdot {}_u\mu''(\zeta)\right] \cdot \frac{h^2}{2}$. The term ${}_u\mu''(\zeta_1) \leq \sup\limits_{u,v} {}_u\mu''(v)$ is bounded by some constant independent of u. The function ${}_u\mu'(v)$ is also bounded due the fact that its derivative is bounded so the term ${}_u\mu'(\zeta)$ is smaller than some constant independent of u. This means $|{}_u m_h^h - {}_u\mu(h)| < Ch^2$, where the constant C does not depend on u and h.

Let $b \geq h$ and $u \in [b, b+h)$ so $u \to b$ as $h \to 0$. By using the Taylor's formula we have

$$\begin{aligned}
{}_u\mu(h) - {}_b\mu(h) &= \frac{\mu(u+h) - \mu(u)}{S(u)} - \frac{\mu(b+h) - \mu(b)}{S(b)} \\
&= \frac{S(b)\mu'(u)h - S(u)\mu'(b)h + O(h^2)}{S(u)S(b)} \\
&= \frac{S(b)\mu'(u) - S(u)\mu'(b)}{S(u)S(b)} \cdot h + O(h^2) = O(h^2)
\end{aligned}$$

and

$$\begin{aligned}
{}_uS(h) - {}_bS(h) &= \frac{S(u+h)S(b) - S(b+h)S(u)}{S(u)S(b)} \\
&= \frac{[S(u+h) - S(u)]S(b) - [S(b+h) - S(b)]S(u)}{S(u)S(b)} \\
&= \frac{S'(u)S(b) - S(u)S'(b) + O(h^2)}{S(u)S(b)} \cdot h = O(h^2).
\end{aligned}$$

This means for $u \in [b, b+h)$ we have

$$\begin{aligned}
{}_u m_h^h &= {}_b\mu(h) + O(h^2), \\
{}_u m_h^{u+h} - {}_u m_h^h &= {}_bS(h) + O(h^2).
\end{aligned}$$

The constant in the error terms $O(h^2)$ does not depend on u and h due to the fact ${}_u\mu''(v)$ and $S''(v)$ are bounded.

Let $\eta_i \in (bh, (b+1)h]$ be the ages of the individuals inside the age interval $(bh, (b+1)h]$ and denote their distribution functions by F_{η_i} for each i. Then,

$$\mathbb{E}\left({}_{\eta_i}z_h^{(b+2)h} - {}_{\eta_i}z_h^{(b+1)h}\right) = \mathbb{E}\left[\mathbb{E}\left({}_{\eta_i}z_h^{(b+2)h} - {}_{\eta_i}z_h^{(b+1)h} \mid \eta_i\right)\right]$$

$$= \int_{bh}^{(b+1)h} \left({}_um_h^{(b+2)h} - {}_um_h^{(b+1)h}\right) \mathrm{d}F_{\eta_i}(u) = \int_{bh}^{(b+1)h} \left({}_um_h^{u+h} - {}_um_h^{h}\right) \mathrm{d}F_{\eta_i}(u)$$

$$= \int_{bh}^{(b+1)h} {}_uS(h)\, \mathrm{d}F_{\eta_i}(u) = {}_{bh}S(h) + O(h^2),$$

$$\mathbb{E}\left({}_{\eta_i}z_h^h\right) = \mathbb{E}\left[\mathbb{E}\left({}_{\eta_i}z_h^h \mid \eta_i\right)\right] = \int_{bh}^{(b+1)h} {}_um_h^h\, \mathrm{d}F_{\eta_i}(u) = {}_{bh}\mu(h) + O(h^2).$$

The number of individuals at age within $(bh; (b+1)h]$ who survived after time h is

$$\mathbb{E}\left(N_{t+h}((b+1)h; (b+2)h]\right) = \mathbb{E} \sum_{i=1}^{N_t(bh;(b+1)h]} \left({}_{\eta_i}z_h^{(b+2)h} - {}_{\eta_i}z_h^{(b+1)h}\right)$$
$$= \mathbb{E}\left(N_t(bh; (b+1)h]\right) \cdot \left({}_{bh}S(h) + O(h^2)\right),$$

and the number of individuals that were born during that time is

$$\mathbb{E}\left(N_{t+h}[0; h)\right) = \sum_{b=0}^{\omega/h-1} \mathbb{E} \sum_{i=1}^{N_t(bh;(b+1)h]} {}_{\eta_i}z_h^h$$
$$= \sum_{b=0}^{\omega/h-1} \left[\mathbb{E}(N_t(bh; (b+1)h]) \cdot \left({}_{bh}\mu(h) + O(h^2)\right)\right]$$
$$= \sum_{b=0}^{\omega/h-1} \left(\mathbb{E}(N_t(bh; (b+1)h]) \cdot {}_{bh}\mu(h)\right) + \mathbb{E}\left(N_t[0; \omega]\right) \cdot O(h^2).$$

If we assume μ to be twice differentiable then

$${}_{bh}\mu(h) = {}_{bh}\mu(0) + {}_{bh}\mu'(0)h + O(h^2) = {}_{bh}\mu'(0)h + O(h^2),$$

from it follows that

$$\mathbb{E}\,(N_{t+h}[0;h)) = \sum_{b=0}^{\omega/h-1} [\mathbb{E}\,(N_t(bh;(b+1)h]) \cdot {}_{bh}\mu'(0)h] + \mathbb{E}\,(N_t[0;\omega]) \cdot O(h^2).$$

□

Let us introduce a matrix notation to lighten the notation:

$$A = \begin{bmatrix} {}_0\mu(h) & {}_h\mu(h) & \dots & {}_{\omega-2h}\mu(h) & {}_{\omega-h}\mu(h) \\ {}_0S(h) & 0 & \dots & 0 & 0 \\ 0 & {}_hS(h) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & {}_{\omega-2h}S(h) & 0 \end{bmatrix}_{\frac{\omega}{h}\times\frac{\omega}{h}},$$

$$B = \begin{bmatrix} {}_0\mu'(0)h & {}_h\mu'(0)h & \dots & {}_{\omega-2h}\mu'(0)h & {}_{\omega-h}\mu'(0)h \\ {}_0S(h) & 0 & \dots & 0 & 0 \\ 0 & {}_hS(h) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & {}_{\omega-2h}S(h) & 0 \end{bmatrix}_{\frac{\omega}{h}\times\frac{\omega}{h}},$$

$$[\mathbb{E}N_t] = \begin{bmatrix} \mathbb{E}\,(N_t[0;h]) \\ \mathbb{E}\,(N_t(h;2h]) \\ \vdots \\ \mathbb{E}\,(N_t(\omega-h;\omega]) \end{bmatrix}_{\frac{\omega}{h}\times 1}, [O(h^2)] = \begin{bmatrix} O(h^2) \\ O(h^2) \\ \vdots \\ O(h^2) \end{bmatrix}_{\frac{\omega}{h}\times 1} \text{ and } [1] = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{\frac{\omega}{h}\times 1}.$$

Corollary 10.1 *Let us substitute $h = 1, k = 1$ in Theorem 10.2 and assume a woman can only have one child in each age interval $(b;(b+1)]$. Let p_b denote the conditional probability of a woman giving birth at certain age within the age interval and s_b denote the conditional probability of a woman surviving to the end of the interval if she is alive at the beginning. Then, the expected number of children that a woman has in the age interval is ${}_b\mu(1) = 1\cdot p_b + 0\cdot(1-p_b) = p_b$, ${}_bS(1) = s_b$, and the matrix A has the following form:*

$$A = \begin{bmatrix} p_0 & p_1 & \dots & p_{\omega-2} & p_{\omega-1} \\ s_1 & 0 & \dots & 0 & 0 \\ 0 & s_2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & s_{\omega-2} & 0 \end{bmatrix}_{\omega\times\omega}.$$

This matrix is also called Leslie matrix in demographics and it is used for projecting human populations (see Keyfitz [6]).

Theorem 10.3 *Let the initial population at time* $t = 0$ *be composed of finite number of individuals on random ages that have absolutely continuous distributions. Let* $\mathbb{E}\,(N_0[0;h])$, $\mathbb{E}\,(N_0(bh;(b+1)h])$, *for* $b = 1, \ldots, (\omega/h - 1)$, *be the expected age structure of the population at time* $t = 0$. *Let the conditions of Theorem 10.1 hold and let* ${}_u\mu''(v)$ *be bounded on* $[0,t]^2$ *and* $S''(v)$ *be bounded on* $[0,t]$. *Then for all* $k \le t/h$ *we have*

$$[\mathbb{E}N_{kh}] = A^k \cdot [\mathbb{E}N_0] + [O(h^2)],$$

$$[\mathbb{E}N_{kh}] = B^k \cdot [\mathbb{E}N_0] + [O(h^2)]$$

and the total population count is

$$[1]^{\mathsf{T}} \cdot [\mathbb{E}N_{kh}] = [1]^{\mathsf{T}} \cdot A^k \cdot [\mathbb{E}N_0] + O(h),$$

$$[1]^{\mathsf{T}} \cdot [\mathbb{E}N_{kh}] = [1]^{\mathsf{T}} \cdot B^k \cdot [\mathbb{E}N_0] + O(h).$$

Proof Theorem 10.2 could be written in matrix form as $[\mathbb{E}N_t] = A\cdot[\mathbb{E}N_{t-h}]+[E_{t-h}]$, where $[E_{t-h}]$ is the error term. If we apply Theorem 10.2 recursively t/h times we obtain

$$[\mathbb{E}N_t] = A^{\frac{t}{h}} \cdot [\mathbb{E}N_0] + A^{\frac{t}{h}-1}[E_0] + \ldots + A[E_{t-2h}] + [E_{t-h}], \tag{10.4}$$

where $[E_{bh}]$ are the error terms.

We will see that the error $[\mathbb{E}N_t]-A^{\frac{t}{h}}\cdot[\mathbb{E}N_0]$ in Eq. (10.4) is actually $[O(h^2)]$. From Theorem 10.2 it follows that each element of the absolute error $[|E_{bh}|]$ is bounded by the corresponding element in the vector $[E_{bh}^{max}]$:

$$[|E_{bh}|] \le \begin{bmatrix} \mathbb{E}\,(N_{bh}[0;\omega]) \cdot Ch^2 \\ \mathbb{E}\,(N_{bh}[0;h]) \cdot Ch^2 \\ \vdots \\ \mathbb{E}\,(N_{bh}(\omega-2h;\omega-h]) \cdot Ch^2 \end{bmatrix}_{\frac{\omega}{h}\times 1} = [E_{bh}^{max}],$$

where the constant C does not depend on the choice of b or h.

From the finiteness of the functions ${}_um'(h)$ and ${}_uS(h)$, it follows that there exists a matrix A_{max} such that each element of the matrices A and B is smaller than the corresponding element in matrix A_{max}:

$$A_{max} = \begin{bmatrix} Mh & Mh & \ldots & Mh & Mh \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}_{\frac{\omega}{h}\times\frac{\omega}{h}},$$

where M is a constant.

Because of the continuous distribution of the point process ξ, the ages of the descendants of each woman in the population are absolutely continuous random variables. If the ages of the initial women at time $t = 0$ are absolutely continuous too, then all the individuals in the population have absolutely continuous random ages for every t, i.e. $\mathbb{E}\left(N_t[b; b+h]\right) \to 0$ for every b and t, when $h \to 0$.

It is obvious that $A \cdot [|E_{bh}|] \leq A_{max} \cdot [E_{bh}^{max}]$, where the comparison is element-wise. If we multiply recursively $A_{max}^{k} \cdot [E_{bh}^{max}] = A \cdot A_{max}^{k-1} \cdot [E_{bh}^{max}]$ and use that

$$\mathbb{E}\left(N_{bh}[0; h]\right) + \ldots + \mathbb{E}\left(N_{bh}(\omega - (k+1)h; \omega - kh]\right) \leq \mathbb{E}\left(N_{bh}[0; \omega]\right),$$

then for $k < \omega/h$,

$$A_{max}^{k} \cdot [E_{bh}^{max}] \leq Ch^2 \cdot \begin{bmatrix} \mathbb{E}\left(N_{bh}[0;\omega]\right) 2Mh(Mh+1)^{k-1} \\ \mathbb{E}\left(N_{bh}[0;\omega]\right) 2Mh(Mh+1)^{k-2} \\ \vdots \\ \mathbb{E}\left(N_{bh}[0;\omega]\right)(2Mh) \\ \mathbb{E}\left(N_{bh}[0;h]\right) \\ \vdots \\ \mathbb{E}\left(N_{bh}(\omega-(k+2)h; \omega-(k+1)h]\right) \end{bmatrix}_{\frac{\omega}{h}\times 1},$$

and for $k \geq \omega/h$

$$A_{max}^{k} \cdot [E_{bh}^{max}] \leq Ch^2 \cdot \begin{bmatrix} \mathbb{E}\left(N_{bh}[0;\omega]\right) 2Mh(Mh+1)^{k-1} \\ \mathbb{E}\left(N_{bh}[0;\omega]\right) 2Mh(Mh+1)^{k-2} \\ \vdots \\ \mathbb{E}\left(N_{bh}[0;\omega]\right) 2Mh(Mh+1)^{k-\omega/h} \end{bmatrix}_{\frac{\omega}{h}\times 1}.$$

Then for $k = t/h$ we have $Mh(Mh+1)^{t/h-1} \sim Me^{tM} \cdot h \to 0$ as $h \to 0$. In addition, $\mathbb{E}\left(N_b[0;\omega]\right)$ is finite for every $b \leq t$ and $\mathbb{E}\left(N_b[a; a+h]\right) \to 0$ for every $a < \omega$ and $b \leq t$, as $h \to 0$. Consequently there exists a constant C, independent of b, and h (but dependent of t) such that each element of vector $A_{max}^{k} \cdot [E_{bh}^{max}]$ is less than Ch^3, $k = 1, \ldots, (t/h - 1)$.

If we sum all error terms in Eq. (10.4) we obtain that every element of the absolute error $|A^{\frac{t}{h}-1}[E_0] + \ldots + A[E_{t-2h}] + [E_{t-h}]|$ is smaller than $\frac{t}{h} \cdot Ch^3 = O(h^2)$ for some constant C hence

$$[\mathbb{E}N_t] = A^{t/h} \cdot [\mathbb{E}N_0] + [O(h^2)].$$

If we sum all of the different ages (their count is ω/h) we obtain

$$[1]^{\mathsf{T}}[\mathbb{E}N_t] = [1]^{\mathsf{T}} \cdot A^{t/h} \cdot [\mathbb{E}N_0] + [1]^{\mathsf{T}} \cdot [O(h^2)] = [1]^{\mathsf{T}} \cdot A^{t/h} \cdot [\mathbb{E}N_0] + O(h).$$

If we fix the period t and decrease h, the error of the formula above tends to 0. However, note that the error term depends on t. If we fix h and increase t, then the approximation error increases. □

Theorem 10.4 *Let $N_0[0; +\infty) = N_0[b; b] = 1$, i.e. we only have one individual at time zero and her age is b. Then, for every $k = 1, \dots, t/h$,*

$$
{}_bm_{kh} = \begin{bmatrix}1 & \dots & 1\end{bmatrix}_{1\times\frac{\omega}{h}} \cdot \begin{bmatrix} {}_0\mu(h) & {}_h\mu(h) & \dots & {}_{\omega-2h}\mu(h) & {}_{\omega-h}\mu(h) \\ {}_0S(h) & 0 & \dots & 0 & 0 \\ 0 & {}_hS(h) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & {}_{\omega-2h}S(h) & 0 \end{bmatrix}^k_{\frac{\omega}{h}\times\frac{\omega}{h}} \cdot \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{\frac{\omega}{h}\times 1} + O(h),
$$

where the only non-zero element in this vector is on position $b/h + 1$, corresponding to the age interval $(b, b + h]$. As alternative we can also use matrix B instead of A for approximation.

Proof In this case the branching process starts with an individual at exact age of b. If we consider a similar branching process starting with an individual of age $\eta \in (b, b + h]$, continuously distributed in the age interval, then we can apply Theorem 10.3. We know that ${}_bm_t$ is a continuous function of b, so $\mathbb{E}_\eta m_t = {}_bm_t + O(h)$. □

Corollary 10.2 *If we consider a branching process starting with an individual of age 0 then by applying Theorem 10.3 we obtain the solutions to the renewal equations (10.1) for all the ages a simultaneously:*

$$
\begin{bmatrix} m^h_{kh} \\ m^{2h}_{kh} - m^h_{kh} \\ \vdots \\ m^\omega_{kh} - m^{\omega-h}_{kh} \end{bmatrix} \approx \begin{bmatrix} {}_0\mu(h) & {}_h\mu(h) & \dots & {}_{\omega-2h}\mu(h) & {}_{\omega-h}\mu(h) \\ {}_0S(h) & 0 & \dots & 0 & 0 \\ 0 & {}_hS(h) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & {}_{\omega-2h}S(h) & 0 \end{bmatrix}^k_{\frac{\omega}{h}\times\frac{\omega}{h}} \cdot \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{\frac{\omega}{h}\times 1}
$$

for every $k = 1, \dots, t/h$. The expected population m^a_{kh} is the sum of the first a/h rows in the column vector above.

Note that if $t \leq \omega$ then we can use a truncated matrix, consisting of the first (t/h) rows and columns and we will obtain the same approximation.

Corollary 10.2 provides us the numerical solution to the renewal equations (10.1). Due to the fact it involves only matrix multiplications and the matrix is actually quite simple (it has non-zero elements only in the first row and one of the diagonals), the numerical method is really fast and efficient. In addition , in every step of the matrix multiplication we obtain the solutions to all the renewal equations at the same time.

Given the initial age structure of the population and the functions $S(t)$ and $\mu(t)$ we can calculate the expected future population age structure by using Theorem 10.2. An example reviewing the population of Bulgaria is presented in the Appendix. Other applications of the GBP in demographics are presented in [11, 12].

Acknowledgements The research is supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No DFNI-I02/17 and partially supported by the financial funds allocated to the Sofia University "St. Kl. Ohridski", grant No 047/2015.

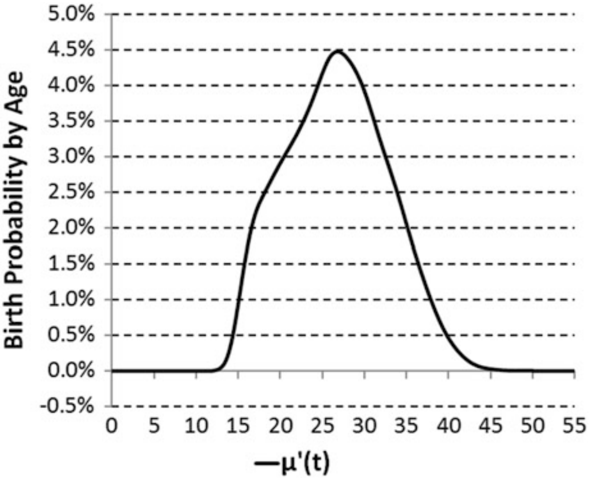
Appendix

Example 1 Forecasting the population age structure of Bulgaria.

In order to find the expected future population of Bulgaria we need to find models for the functions $S(t)$ and $\mu(t)$ from empirical data. The data used for calculations in this paper are published by Eurostat. One knows the number of live births by mother's age, the number of deaths by age and the number of people by age. The demographic theory uses these data to produce the life-tables (see Keyfitz [6]). The Chiang's formula (see Chiang [3]) presents the relation between age-specific birth and death rates and the probabilities for birth and death by age. Using demographics we can find the distribution of the life length and the probability of a woman giving birth at a specific age from the demographic coefficients.

The empirical estimations of these age-specific probabilities suggest the assumption of smoothness for $\mu(t)$ and $S(t)$ is appropriate in case of human populations. Using smoothing splines, we can find the functional model that generated the data (see Ramsay [10] and Boor [2]). The resulting functions for $\mu(t)$ and $S(t)$ are presented in Figs. 10.1, 10.2, and 10.3.

Fig. 10.1 Point process density function $\mu'(t)$, Bulgaria 2013



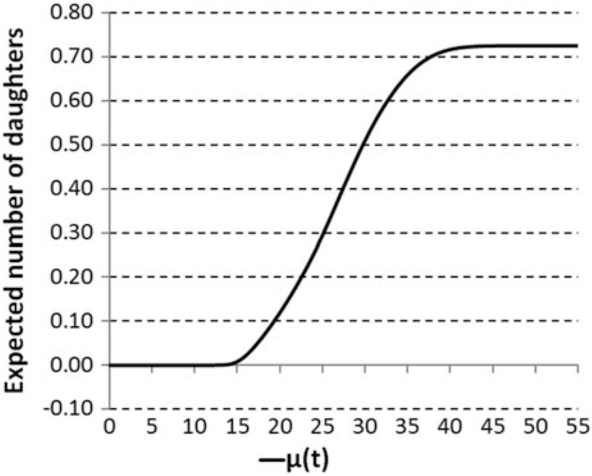


Fig. 10.2 Expectation of point process $\mu(t)$, Bulgaria 2013

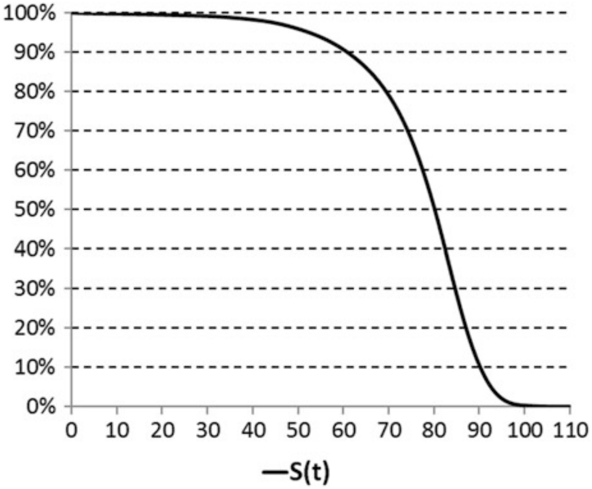


Fig. 10.3 Survival probability $S(t)$, Bulgaria 2013

We know the initial age structure of the population $N_0(bh; (b+1)h]$, i.e. the number of all women on a specific age at the beginning of 2013. Having modelled the life length of each individual and her point process from real data (published by Eurostat), we can substitute those in Corollary 10.2 and Theorem 10.2. On each step of the matrix multiplication, we obtain the expected future population age structure. The resulting total population count is presented in Fig. 10.4.

Example 2 Numerical solution to renewal equations.

Let us consider a Poisson renewal process, i.e. the inter-renewal time is exponentially distributed according to the distribution function $F(x) = 1 - e^{-\lambda x}$. It is well-known that the number of renewals until time t follows Poisson distribution and the expected number of renewals until time t is given by $\mathbb{E}N(t) = \lambda t$. The

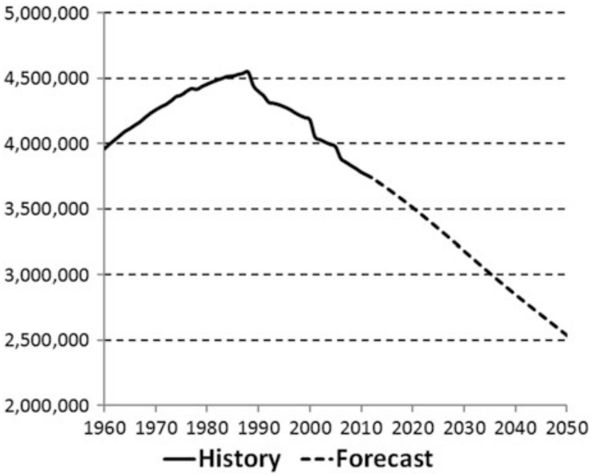


Fig. 10.4 Female population count

function $U(t) = \mathbb{E}N(t) + 1 = 1 + \lambda t$ is called renewal function and it satisfies the following renewal equation.

$$U(t) = I(t) + \int_0^t U(t-u)dF(u), \tag{10.5}$$

where $I(t) = 1$ for $t \geq 0$. If we consider the general equation

$$Z(t) = z(t) + \int_0^t Z(t-u)dF(u), \tag{10.6}$$

then (from renewal theory) its solution is $Z(t) = (U * z)(t)$. To see how the numerical method in Corollary 10.2 works we will use it for solving Eq. (10.5) and then Eq. (10.6), choosing for example $z(t) = e^{-t}$.

From Eq. (10.5) it is obvious that $S(t) = 1$ for every $t > 0$ and

$$_{bh}\mu(h) = \frac{\mu((b+1)h) - \mu(bh)}{S(bh)} = e^{-\lambda bh}(1 - e^{-\lambda h})$$

for every b. From $_b\mu(t) = [\mu(t+b) - \mu(b)]/S(b)$ we can calculate the derivative $_b\mu'(t) = \mu'(t+b)/S(b)$. Then $_{bh}\mu'(0) = \mu'(bh)/S(bh) = \mu'(bh) = \lambda e^{-\lambda bh}$.

In this case $\omega = +\infty$ but for the purpose of calculating the $U(t)$ we can assume $\omega = t$ or we can assign some value that is greater than t to ω and the results will not change. However, it is more convenient for calculations to use the smallest possible

$\omega = t$ so that the matrix A has the following form:

$$A = \begin{bmatrix} (1-e^{-\lambda h}) & \frac{(1-e^{-\lambda h})}{e^{-\lambda h}} & \dots & \frac{(1-e^{-\lambda h})}{e^{-\lambda(t-2h)}} & \frac{(1-e^{-\lambda h})}{e^{-\lambda(t-h)}} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Let us denote by $\hat{U}(t)$ the numerical solution of Eq. (10.5). It can be shown that $A^k \cdot \left[1\ 0 \dots 0\right]^\mathsf{T} \approx \left[\lambda h \dots \lambda h\ 1\ 0 \dots 0\right]^\mathsf{T}$ and $\hat{U}(kh) = [1] \cdot A^k \cdot \left[1\ 0 \dots 0\right]^\mathsf{T} = 1 + (kh)\lambda + k \cdot O(h^2) = U(kh) + k \cdot O(h^2)$ for every $k = 1, \dots, t/h$, using Taylor expansion of e^x. For $k = t/h$, we obtain $U(t) = \hat{U}(t) + O(h)$.

Let us consider the second equation (10.6). Its theoretical solution is given by the Riemann-Stieltjes integral

$$\begin{aligned} Z(t) &= \int_0^t e^{-(t-u)} d(1+\lambda u) = e^{-t} + \lambda \int_0^t e^{u-t} d(u-t) \\ &= e^{-t} + \lambda e^{u-t}|_0^t = e^{-t} + \lambda(1-e^{-t}). \end{aligned}$$

Note that the function U has a jump for $t = 0$: $U(0) = 1$ and $U(t) = 0$ for $t < 0$, so the term $U(0) \cdot z(t-0) = e^{-t}$ is added when calculating the Stieltjes integral.

In this particular case, we have $S(t) = e^{-t}$, ${}_{bh}S(h) = S((b+1)h)/S(bh) = e^{-h}$ and ${}_{bh}\mu(h) = [\mu((b+1)h) - \mu(bh)]/S(bh) = e^{-\lambda bh}(1-e^{-\lambda h})/e^{-bh}$ for every b. Then,

$$A = \begin{bmatrix} (1-e^{-\lambda h}) & \frac{e^{-\lambda h}(1-e^{-\lambda h})}{e^{-h}} & \dots & \frac{e^{-\lambda(t-2h)}(1-e^{-\lambda h})}{e^{-(t-2h)}} & \frac{e^{-\lambda(t-h)}(1-e^{-\lambda h})}{e^{-(t-h)}} \\ e^{-h} & 0 & \dots & 0 & 0 \\ 0 & e^{-h} & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & e^{-h} & 0 \end{bmatrix}.$$

Since it can be shown by recursion that

$$A^k \left[1\ 0 \dots 0\right]^\mathsf{T} \approx \left[\lambda h \dots \lambda h\ (1-kh)\ 0 \dots 0\right]^\mathsf{T},$$

the numerical approximation of the function $Z(t)$ is $\hat{Z}(kh) = [1] \cdot A^k \cdot \left[1\ 0 \dots 0\right]^\mathsf{T} = 1 + (kh)\lambda - kh + k \cdot O(h^2)$ for every $k = 1, \dots, t/h$. The theoretical solution is $Z(kh) = e^{-kh} + \lambda(1-e^{-kh}) = 1 - kh + \lambda \cdot kh + O(h^2)$ consequently it follows that $U(kh) = \hat{U}(kh) + k \cdot O(h^2)$. When $k = t/h$ we get $U(t) = \hat{U}(t) + O(h)$. For a fixed

t, we obtain the numerical solution of the renewal equation inside the interval $[0, t]$ with estimation error $O(h)$.

References

1. Bartholomew, D.: An approximate solution to the integral equation of renewal theory. J. R. Statist. Soc. B **25**, 432–441 (1963)
2. Boor, C.: A Practical Guide to Splines. Springer, New York (2001)
3. Chiang, C.: Life Table and Its Applications. Krieger Publishing Co., Malabar, FL (1983)
4. From, S.: Some new approximations for the renewal function. Commun. Stat. Simul. Comput. **30**(1), 113–128 (2001)
5. Jagers, P.: Branching Processes with Biological Applications. Wiley, London/New York/Sydney/Toronto (1975). http://onlinelibrary.wiley.com/doi/10.1002/bimj.4710200112/abstract
6. Keyfitz, N., Caswell, H.: Applied Mathematical Demography. Springer, New York (2005)
7. Mode, C.: Stochastic Processes in Demography and Their Computer Implementation. Springer, Berlin/Heidelberg (1985). http://www.springer.com/us/book/9783642823244
8. Mitov, K., Omey, E.: Renewal Processes. Springer, New York (2013)
9. Mitov, K., Omey, E.: Intuitive approximations for the renewal function. Stat. Probab. Lett. **84**, 72–80 (2014)
10. Ramsay, J., Silverman, B.: Functional Data Analysis. Springer, New York (2005)
11. Trayanov, P., Slavtchova-Bojkova, M.: Crump-Mode-Jagers branching processes: application in population projections. Advanced research in mathematics and computer science. In: Proceedings of the Doctoral Conference in Mathematics, Informatics and Education, Sofia, pp. 1–8, 19–20 September 2013
12. Trayanov, P., Slavtchova-Bojkova, M.: Estimating the effect of economic crisis with Crump-Mode-Jagers branching process. Pliska Stud. Math. Bulgar. **24**, 99–110 (2015)
13. Xie, M.: On the solution of renewal-type integral equations. Commun. Stat. Simul. Comput. **18**(1), 281–293 (1989)

Part VI
Special Branching Models

Chapter 11
Bayesian Analysis for Controlled Branching Processes

Miguel González, Cristina Gutiérrez, Rodrigo Martínez, Carmen Minuesa, and Inés M. del Puerto

Mathematics Subject Classification (2000): 60J80

11.1 Introduction

The branching model considered in the present work is the controlled branching process. This model is a generalization of the standard Bienaymé-Galton-Watson (BGW) branching process, and, in the terminology of population dynamics, is used to describe the evolution of populations in which a control of the population size at each generation is needed. This control consists of determining mathematically the number of individuals with reproductive capacity at each generation through a random process. In practice, this branching model can describe reasonably well the probabilistic evolution of populations in which, for various reasons of an environmental, social, or other nature, there is a mechanism that establishes the number of progenitors which take part in each generation. For example, in an ecological context, one can think of an invasive animal species that is widely recognized as a threat to native ecosystems, but there is disagreement about plans to eradicate it, i.e., while the presence of the species is appreciated by a part of the society, if its numbers are left uncontrolled it is known to be very harmful to native ecosystems. In such a case, it is better to control the population to keep it within

---

M. González (✉) • C. Gutiérrez • C. Minuesa • I.M. del Puerto
Department of Mathematics, University of Extremadura, Avda. Elvas, 06006 Badajoz, Spain
e-mail: mvelasco@unex.es; cgutierrez@unex.es; cminuesaa@unex.es; idelpuerto@unex.es

R. Martínez
Centro Universitario de Plasencia, Department of Mathematics, University of Extremadura, Plasencia, Spain
e-mail: rmartinez@unex.es

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_11

admissible limits even though this might mean periods when animals have to be culled. Another practical situation that can be modelled by this kind of process is the evolution of an animal population that is threatened by the existence of predators. In each generation, the survival of each animal (and therefore the possibility of giving new births) will be strongly affected by this factor, making the introduction of a random mechanism necessary to model the evolution of this kind of population.

Mathematically, a controlled branching process with random control function (CBP) is a discrete-time stochastic growth population model $\{Z_n\}_{n\geq 0}$ defined recursively as

$$Z_0 = N \in \mathbb{N}, \quad Z_{n+1} = \sum_{j=1}^{\phi_n(Z_n)} X_{nj}, \quad n \geq 0, \tag{11.1}$$

where $\{X_{nj} : n = 0, 1, \ldots; j = 1, 2, \ldots\}$ and $\{\phi_n(k) : n, k = 0, 1, \ldots\}$ are two independent families of non-negative integer-valued random variables. Moreover, X_{nj}, $n = 0, 1, \ldots; j = 1, 2, \ldots$, are independent and identically distributed (i.i.d.) random variables and, for each $n = 0, 1, \ldots$, $\{\phi_n(k)\}_{k\geq 0}$, are independent stochastic processes with equal one-dimensional probability distributions. The empty sum in (11.1) is defined to be 0. Let $\{p_k : k \geq 0\}$ denote the common probability distribution of the random variables X_{nj}, i.e., $p_k = P(X_{nj} = k)$, $k \geq 0$, and $m = E[X_{nj}]$ (assumed finite).

Intuitively, Z_n denotes the number of individuals in generation n, and X_{ni} the number of offspring of the ith individual in generation n. Thus, the probability law $\{p_k : k \geq 0\}$ is termed the offspring distribution, and m the offspring mean. The control variables $\phi_n(\cdot)$ could be seen as a random mechanism determining the individual migration process in each generation depending on its population size.

The probabilistic theory of CBPs, in particular the problem of their extinction and their limiting behaviour, has been extensively investigated (see for example [1, 8] and references therein, and [18]). The presence of the control mechanism complicates the study of this kind of process. Nevertheless, it allows one to model a much greater variety of behaviours than the BGW branching process. One of its features that is important for applications is that it overcomes one of the main deficiencies of the BGW process, which is the possibility of an unbounded population growth with a supercritical offspring law (i.e., $m > 1$). Thus, in [5], it was proved that a CBP with offspring mean greater than unity (and, of course, other regularity conditions governing the control mechanism) can die out with probability one. The offspring mean continues to play a key role in the probabilistic evolution of a CBP, as the above-cited papers show. Indeed, it was established that the asymptotic mean growth rate of the process denoted by τ and defined by $\tau = \lim_{k\to\infty} k^{-1}E[Z_{n+1} \mid Z_n = k] = \lim_{k\to\infty} k^{-1}\varepsilon(k)m$, with $\varepsilon(k) = E[\phi_n(k)]$ (whenever it exists) is the threshold parameter that determines the behaviour of a CBP in relation to its extinction. Hence the importance of making inferences on the offspring mean and on the asymptotic mean growth rate. However, there have as yet been few papers devoted to this topic. Inferential studies from a frequentist standpoint may be found in [6, 7, 11, 19]. A first approach from a Bayesian standpoint was considered in [13] in a parametric context, and in [10] for the particular case of a deterministic control function.

The present study is a continuation of this line of research by developing the inferential theory in a non-parametric framework for the offspring law and in a parametric setting for the control distributions, depending on a single parameter termed the control parameter. Notice that, in relation to [10], random control rather than deterministic one introduces much more uncertainty in the model becoming its behaviour richer, but adding a considerable difficulty on the estimation of its parameters. Moreover, another important novelty is that we assume an upper bound of the support of the offspring law is unknown. We model this uncertainty considering the support potentially infinite. Although this could seem a strong condition, the gain in flexibility makes the model more attractive and realistic. In this case, to deal with the inference procedure we shall use as prior a Dirichlet process on the space of the nonnegative integers. Consequently, in this chapter, we address the inference of the control parameter, of the offspring distribution, and of the offspring mean, as well as the asymptotic mean growth rate and the prediction of future sizes of the population. To this end, Sect. 11.2 begins by assuming that the entire family tree up to some given generation can be observed. A Dirichlet process is introduced to model the prior distribution of the offspring law, avoiding assumptions on the cardinality of its support. However, actually, in most populations, it is not possible to observe these data, and only the population size at each generation can be recorded. To deal with the Bayesian inference in this case, a Markov chain Monte Carlo (MCMC) method is used, in particular, the Gibbs sampler algorithm, to approximate the posterior distribution of parameters of interest. The present implementation of this algorithm generalizes the results in [9, 10] and represents the Bayesian analogy of the given in [11].

Section 11.2 also deals with the problem of approximating the predictive posterior distributions. As illustration, in Sect. 11.3 a simulated example is presented. Finally, some concluding remarks are given in Sect. 11.4.

11.2 Bayesian Analysis

For the purpose of this chapter, we consider a CBP with an offspring distribution $p = \{p_k : k \geq 0\}$, without assuming any knowledge about the cardinality of its support. With respect to the random control mechanism, note that one has different probability distributions for each population size $k \geq 0$, that corresponding to $\phi_n(k)$. Consequently, from a finite sample, it is not possible to deal with the inference problems arising from this model (at least for the control distributions) without assuming that there exists some stable structure. We therefore consider a parametric scheme for the control process. In particular, we take a CBP with control distributions belonging to the power series family. Formally, for each $k \geq 0$,

$$P(\phi_n(k) = j) = a_k(j)\theta^j / A_k(\theta), \quad j = 0, 1, \ldots ; \theta \in \Theta_k, \tag{11.2}$$

with $a_k(j)$ known non-negative values, $A_k(\theta) = \sum_{j=0}^{\infty} a_k(j)\theta^j$, and $\Theta_k = \{\theta > 0 : 0 < A_k(\theta) < \infty\}$ being an open subset of $\mathbb{R}$. We also assume that the sets Θ_k are independent of k. Hence, we shall henceforth drop the index k from Θ_k, the control parameter space. This implies a certain stability in the probability laws of the control variables $\phi_n(k)$, for $k \geq 0$, depending on a single parameter θ, the control parameter, and on the size of the population, k. Let us write $\epsilon(k, \theta) = E[\phi_n(k)]$, and assume the following regularity condition:

$$\prod_{k \in B} A_k(\theta) = A_{\sum_{k \in B} k}(\theta), \text{ for every } B \subseteq \mathbb{N},\ \theta \in \Theta. \tag{11.3}$$

Condition (11.3), satisfied by a wide family of probability distributions, is a technical hypothesis that allows the theory of conjugate families to be made use of in the Bayesian analysis to be developed below. Moreover (11.3) implies that $A_k(\theta) = (A_1(\theta))^k$, so that

$$\epsilon(k, \theta) = \theta A_k'(\theta)(A_k(\theta))^{-1} = k\theta(A_1(\theta))^{k-1} A_1'(\theta)(A_1(\theta))^{-k} = k\epsilon(1, \theta).$$

This allows the family of distributions verifying (11.3) to be reparametrized by $\mu = \mu(\theta) = \epsilon(1, \theta)$. It also guarantees the existence of the limit τ and its parametrization in terms of the two mean-value parameters, the offspring mean m and μ, with $\tau = m\mu$.

Unlike the parameter θ, μ has the same interpretation for all the power series families of distributions satisfying (11.3). We refer to μ as the migration parameter since, whenever $\mu < 1$, one can use such control distributions to model processes with expected emigration, whereas $\mu > 1$ can model processes with expected immigration. For $\mu = 1$ neither emigration nor immigration is expected.

Remark 11.1 Some interesting particular cases of control distributions verifying (11.2) and (11.3) are the following:

a) Consider for each k that $\phi_n(k)$ follows a binomial distribution with parameters k and $0 < q < 1$. In such a case $\theta = q(1 - q)^{-1}$. It is easy to see that conditions (11.2) and (11.3) hold, and in particular that $\mu = \theta(1 + \theta)^{-1}(= q)$. From a practical standpoint, this control mechanism could be reasonable to model situations in which, in each generation, each individual can be removed from the population with probability $1 - q$, not participating in its subsequent evolution, and can survive and give birth to offspring in the next generation with probability q. The value of q is considered to be independent of the total population size in each generation. These control functions always take into account the possibility of an emigration phenomenon in each generation ($\mu < 1$). A CBP with this family of control distributions can be useful to model predator effects in a population.
b) Consider for each k that $\phi_n(k)$ has a Poisson distribution with parameter $k\theta$. Again conditions (11.2) and (11.3) hold, and one deduces that $\mu = \theta$. Depending

on the value of μ, a CBP with these control functions can model either expected immigration or expected emigration processes.

c) Consider for each k that $\phi_n(k)$ has a negative binomial distribution with parameters k and q, with $0 < q < 1$. Now $\theta = 1 - q$. It is easy to see again that conditions (11.2) and (11.3) hold. In this case $\mu = \theta(1-\theta)^{-1}$. The associated CBP can also be used to model different migratory movements.

To infer the posterior distribution of the main parameters of interest, we consider initially that the entire family tree up to the current nth generation can be observed. After studying this case, we consider a more realistic requirement that only the total population size at each generation can be sampled. To deal with this second situation, we shall need the Gibbs sampler.

11.2.1 Analysis Based on the Entire Family Tree

We consider that the entire family tree up to the current nth generation can be observed, i.e., $\{X_{lj} : j = 1, \ldots, \phi_l(Z_l);\ l = 0, 1, \ldots, n-1\}$, or at least the variables

$$Z^*_{n,k} = \sum_{l=0}^{n-1} Z_l(k), \text{ where } Z_l(k) = \sum_{j=1}^{\phi_l(Z_l)} I_{\{X_{lj}=k\}},\ k \geq 0,$$

with I_A standing for the indicator function of the set A. Intuitively, $Z_l(k)$ represents the number of progenitors at the lth generation with exactly k offspring, and therefore $Z^*_{n,k}$ is the accumulated number up to generation n of progenitors that give rise to exactly k offspring. Let us write $\mathcal{Z}^*_n = \{Z_l(k),\ k \geq 0,\ l = 0, 1, \ldots, n-1\}$, and introduce the following variables:

$$Y_n = \sum_{l=0}^{n-1} Z_l \quad \text{and} \quad Y^*_n = \sum_{l=0}^{n-1} \phi_l(Z_l),$$

i.e., Y_n and Y^*_n represent, respectively, the total number of individuals and progenitors in the population up to the $(n-1)$th generation. Using (11.2) and (11.3), one can deduce that the likelihood based on the sample $\mathcal{Z}^*_n$ verifies

$$f(\mathcal{Z}^*_n \mid p, \theta) \propto \prod_{k \geq 0} p_k^{Z^*_{n,k}} \theta^{Y^*_n} / A_{Y_n}(\theta). \tag{11.4}$$

Hence, since no restriction has been imposed on the cardinality of support of the reproduction law, and the offspring and control distributions are independent, an appropriate conjugate class of prior distributions for (p, θ) is $\pi(p, \theta) = \pi(p)\pi(\theta)$, with $\pi(p)$ being the distribution corresponding to

$$p \sim \mathrm{DP}(p(0), \alpha), \tag{11.5}$$

where DP denotes the Dirichlet process, with $p(0) = \{p_k(0),\ k \geq 0\}$ being the base measure and α the concentration parameter, $\alpha > 0$, and $\pi(\theta)$ the distribution given by the density

$$\varphi(a,b)^{-1}\theta^a / A_b(\theta), \tag{11.6}$$

with

$$\varphi(a,b) = \int_\Theta \theta^a / A_b(\theta) d\theta,$$

where $a, b \geq 0$.

Then, using (11.4)–(11.6), one has that the posterior distribution

$$\pi(p, \theta \mid \mathcal{Z}_n^*) \propto \pi(p \mid \mathcal{Z}_n^*) \pi(\theta \mid \mathcal{Z}_n^*), \tag{11.7}$$

with $\pi(p \mid \mathcal{Z}_n^*)$ being the distribution corresponding to

$$p \mid \mathcal{Z}_n^* \sim \mathrm{DP}\left(\frac{\alpha}{\alpha + Y_n^*} p(0) + \frac{1}{\alpha + Y_n^*} \sum_{k \geq 0} Z_{n,k}^* \delta_k, \alpha + Y_n^*\right),$$

with δ_k a Dirac delta at k, $k \geq 0$, and

$$\pi(\theta \mid \mathcal{Z}_n^*) = \varphi(a + Y_n^*, b + Y_n)^{-1} \theta^{a + Y_n^*} / A_{b + Y_n}(\theta).$$

From (11.7), using Dirichlet process properties and considering the squared error loss function, it follows straightforwardly that the Bayes estimator for the offspring distribution and θ are, respectively:

$$\widehat{p}_k = (\alpha p_k(0) + Z_{n,k}^*) / (\alpha + Y_n^*), \quad k \geq 0,$$

and

$$\hat{\theta} = \varphi(a + Y_n^* + 1, b + Y_n) / \varphi(a + Y_n^*, b + Y_n).$$

As a consequence, one obtains that the Bayes estimator for the offspring mean based on the sample $\mathcal{Z}_n^*$, under squared error loss, is given by

$$\tilde{m} = (\alpha m^{(0)} + Y_n + Z_n - Z_0) / (\alpha + Y_n^*), \tag{11.8}$$

with $m^{(0)}$ being the mean of $p(0)$, and for μ and τ one has $\tilde{\mu} = \int_{\Theta} \mu(\theta)\pi(\theta \mid \mathcal{Z}_n^*)d\theta$ and $\tilde{\tau} = \tilde{m}\tilde{\mu}$, respectively.

Remark 11.2 In particular, for the examples considered in Remark 11.1, using (11.6), one has that the beta distribution family is an appropriate conjugate class of prior distributions for q in the binomial control case and for θ when the control distributions are negative binomial. For the Poisson control distribution case, an appropriate conjugate class of priors for θ is the gamma distribution family.

Remark 11.3 The theoretical approach to dealing with the inference issues related to a CBP is to assume that the control law belongs to the power series distribution family. This is an exponential family that includes many important distributions. It is worth noting that, from a practical standpoint, in most situations the choice of the control process, whether it is governed by a Poisson, binomial, negative binomial, or some other scheme, should be a prior specification based on knowledge of the development of the population.

11.2.2 Analysis Based on Population Size in Each Generation: Gibbs Sampler

In real situations, it is difficult to observe the whole family tree up to the current generation or even the random variables $Z_l(k)$, $k \geq 0$, $l = 0, \ldots, n-1$. Hence, in this subsection we shall assume the more realistic requirement that these are unobservable, with the observable data being $\mathcal{Z}_n = \{Z_0, \ldots, Z_n\}$. Given the definition of the model, an expression of the posterior distribution for (p, θ) after observing $\mathcal{Z}_n$ can not be displayed in a closed form. Consequently, we shall describe an algorithm based on the Gibbs sampler (see e.g., [2]) to approximate it only by observing $\mathcal{Z}_n$. To this end, it is necessary to take the unobservable variables $Z_l(k)$, $k \geq 0$, $l = 0, 1, \ldots, n-1$ as being latent variables, and consider the augmented parameter vector $(p, \theta, \mathcal{Z}_n^*)$. Let $\pi(p, \theta \mid \mathcal{Z}_n)$ denote the posterior distribution of (p, θ) after observing $\mathcal{Z}_n$. We shall approximate the posterior distribution of $(p, \theta, \mathcal{Z}_n^*)$ after observing $\mathcal{Z}_n$, denoted by $\pi(p, \theta, \mathcal{Z}_n^* \mid \mathcal{Z}_n)$, and from this obtain an approximation for its marginal distribution $\pi(p, \theta \mid \mathcal{Z}_n)$. To use the Gibbs sampler, first, it is necessary to obtain the conditional posterior distribution of (p, θ) after observing $\mathcal{Z}_n$ and $\mathcal{Z}_n^*$, which is denoted by $\pi(p, \theta \mid \mathcal{Z}_n, \mathcal{Z}_n^*)$, and the conditional posterior distribution of $\mathcal{Z}_n^*$ after observing $(p, \theta, \mathcal{Z}_n)$, denoted by $f(\mathcal{Z}_n^* \mid p, \theta, \mathcal{Z}_n)$.

Taking into account that, for $l = 0, \ldots, n-1$,

$$Z_{l+1} = \sum_{k \geq 0} k Z_l(k), \tag{11.9}$$

$\pi(p, \theta \mid \mathcal{Z}_n, \mathcal{Z}_n^*)$ is the same as $\pi(p, \theta \mid \mathcal{Z}_n^*)$ given in (11.7). Let us now consider $f(\mathcal{Z}_n^* \mid p, \theta, \mathcal{Z}_n)$. Denoting by $P(\cdot)$ the conditional probability given an offspring distribution p and control distributions governed by θ (the explicit indication of the conditioning on p and θ is dropped for notational clarity), since the individuals reproduce independently and the control distributions are independent of the offspring distribution, one has that, for $z_l(k) \in \mathbb{N} \cup \{0\}$, $k \geq 0$, $l = 0, 1, \ldots, n-1$, $z_l \in \mathbb{N}$, $l = 0, \ldots, n$, satisfying the constraints $z_l = \sum_{k\geq 0} k z_{l-1}(k)$, $l = 1, \ldots, n$,

$$\begin{aligned} &P(Z_l(k) = z_l(k),\ k \geq 0,\ l = 0, 1, \ldots, n-1 \mid Z_0 = z_0, \ldots, Z_n = z_n) \\ &\quad = \prod_{l=0}^{n-1} P\left(Z_l(k) = z_l(k),\ k \geq 0 \mid Z_l = z_l, Z_{l+1} = z_{l+1}\right). \end{aligned}$$

Hence,

$$f(\mathcal{Z}_n^* \mid p, \theta, \mathcal{Z}_n) = \prod_{l=0}^{n-1} f(Z_l(k), k \geq 0 | p, \theta, Z_l, Z_{l+1}),$$

where $f(Z_l(k), k \geq 0 | p, \theta, Z_l, Z_{l+1})$ denotes the conditional distribution of the random sequence $\{Z_l(k), k \geq 0\}$ given p, θ, Z_l, and Z_{l+1}. Now, writing $\phi_l^* = \sum_{k\geq 0} z_l(k)$,

$$\begin{aligned} &P(Z_l(k) = z_l(k),\ k \geq 0 \mid Z_l = z_l, Z_{l+1} = z_{l+1}) \\ &\quad = \frac{1}{P(Z_{l+1} = z_{l+1} \mid Z_l = z_l)} \frac{\phi_l^*!}{\prod_{k\geq 0} z_l(k)!} \prod_{k\geq 0} p_k^{z_l(k)} a_{z_l}\left(\phi_l^*\right) \theta^{\phi_l^*} / A_{z_l}(\theta). \end{aligned}$$

Thus, computationally, an appropriate way to obtain a sample from $f(\mathcal{Z}_n^* \mid p, \theta, \mathcal{Z}_n)$ is as follows. Given the known sample $\{z_0, \ldots, z_n\}$ and known values of θ and p, one samples, for each $l = 0, 1, \ldots, n-1$, a value $\phi_l^*(z_l)$ from the distribution of the variable $\phi_l(z_l)$ given by (11.2). Then, for each $l = 0, 1, \ldots, n-1$, one samples a sequence $\{z_l(k),\ k \geq 0\}$ from the multinomial probabilities $\frac{\phi_l^*(z_l)!}{\prod_{k\geq 0} z_l(k)!} \prod_{k\geq 0} p_k^{z_l(k)}$, $k \geq 0$, normalized by considering the constraint $z_{l+1} = \sum_{k\geq 0} k z_l(k)$. Notice that, although the cardinality of the support of the reproduction law may be infinite, once z_{l+1} is known, only a finite number of coordinates of the sequence $\{z_l(k), k \geq 0\}$ are non-null. Indeed, $z_l(k) = 0$ for all $k \geq z_{l+1}$.

Once it is known how to obtain samples from the distributions $\pi(p, \theta \mid \mathcal{Z}_n, \mathcal{Z}_n^*)$ and $f(\mathcal{Z}_n^* \mid p, \theta, \mathcal{Z}_n)$, the Gibbs sampler algorithm works as follows:

```
Initialize $l = 0$
Generate $p^{(0)} \sim \mathtt{DP}(p(0), \alpha)$
Generate $\theta^{(0)}$ from (11.6)
Iterate
   $l = l + 1$
   Generate $\mathcal{Z}_n^{*(l)} \sim f(\mathcal{Z}_n^* \mid p^{(l-1)}, \theta^{(l-1)}, \mathcal{Z}_n)$
   Generate $(p^{(l)}, \theta^{(l)}) \sim \pi(p, \theta \mid \mathcal{Z}_n^{*(l)})$
```

Notice that, given the sample $\mathcal{Z}_n$, the maximum number of coordinates of $p^{(l)}$, for all $l \geq 0$, involved in the algorithm is $1+\max_{1\leq k\leq n}\{Z_k\}$. Hence, in the last step of the algorithm, bearing in mind Eq. (11.7) and the properties of the Dirichlet process, one obtains these probabilities from the Dirichlet distribution.

The sequence $\{(p^{(l)}, \theta^{(l)}, \mathcal{Z}_n^{*(l)})\}_{l\geq 0}$ is an ergodic Markov chain, and the stationary distribution of that Markov chain is just the sought-after joint distribution, $\pi(p, \theta, \mathcal{Z}_n^* \mid \mathcal{Z}_n)$. Several practical implementation issues must be taken into account for success with the sample obtained by the method described above. Common approaches to reaching the equilibrium distribution as well as to reducing the autocorrelation in the sample are to choose a sufficient burn-in period, N, and to thin the output by storing only every Gth value after the burn-in period (G is known as the batch size). Thus, for a run of the sequence $\{(p^{(l)}, \theta^{(l)}, \mathcal{Z}_n^{*(l)})\}_{l\geq 0}$, one chooses $Q+1$ vectors $\{(p^{(N)}, \theta^{(N)}), (p^{(N+G)}, \theta^{(N+G)}), \ldots, (p^{(N+QG)}, \theta^{(N+QG)})\}$. These vectors are approximately independent sampled values of the distribution $\pi(p, \theta|\mathcal{Z}_n)$ if G and N are large enough (see [20]). Since they could be affected by the initial state $(p^{(0)}, \theta^{(0)})$, the algorithm is applied T times, obtaining a final sample of length $T(Q+1)$. To determine N, G, and T in practice, we shall make use of the Gelman–Rubin–Brooks and autocorrelation diagnostics (see [3, 4]). From this sample one can estimate $\pi(p, \theta \mid \mathcal{Z}_n)$ and its marginal distributions, $\pi(p \mid \mathcal{Z}_n)$ and $\pi(\theta \mid \mathcal{Z}_n)$, by making use of kernel density estimators. These posterior densities can be used to calculate numerically highest-probability-density (HPD) credible sets for the respective parameters, yielding sets in which there is a high probability of finding those parameters. In general, if $\Psi(p, \theta)$ denotes a function of the offspring law and the control parameter (we shall be interested below in m, μ, and τ) then

$$\pi(\Psi \mid \mathcal{Z}_n) = \int \pi(\Psi \mid \mathcal{Z}_n, p, \theta)\pi(p, \theta \mid \mathcal{Z}_n)dpd\theta.$$

Again using kernel density estimators, one can also approximate $\pi(\Psi \mid \mathcal{Z}_n)$ and calculate its HPD sets.

11.2.3 Approaches to Prediction

A very important problem from a practical standpoint is to infer the size of future generations from currently available information. Thus, from a sample $\{Z_0, \ldots, Z_n\}$, one desires inferential statements about unobserved Z_{n+l}, $l \geq 1$. Few results related to this topic can be found in the branching process theory literature (see [12, 14] for BGW processes). From a Bayesian standpoint, any inferential statement about $Z_{n+l}, l \geq 1$, given known population sizes until generation n, is contained in the posterior predictive distribution $f(Z_{n+l} \mid \mathcal{Z}_n)$. Of course, the inferential content of the predictive distribution may be appropriately summarized to provide an estimator of Z_{n+l} as the mean of $f(Z_{n+l} \mid \mathcal{Z}_n)$, and interval estimates of Z_{n+l} such as the class of HPD sets which may be derived from $f(Z_{n+l} \mid \mathcal{Z}_n)$. For simplicity, we shall focus on the set $\{Z_{n+l} > 0\}$, avoiding approximating the density of the mass point $\{Z_{n+l} = 0\}$ whose estimation is obvious. The difficulty lies in finding a closed form for this distribution.

We shall present two ways of approximating the predictive distribution, both applying a Monte Carlo procedure. The first is a sampling-based method, and the second is based on approximating $E[Z_{n+l} \mid \mathcal{Z}_n]$ and $Var[Z_{n+l} \mid \mathcal{Z}_n]$ and then seeking a parametric model.

Method A We consider a random sample $\{(p^{(1)}, \theta^{(1)}), \ldots, (p^{(r)}, \theta^{(r)})\}$ from $\pi(p, \theta \mid \mathcal{Z}_n)$. In particular, we shall use the one obtained with the Gibbs sampler in Sect. 11.2.2, i.e., $r = T(Q + 1)$. It is clear that

$$f(Z_{n+l} \mid \mathcal{Z}_n) = \int f(Z_{n+l} \mid \mathcal{Z}_n, p, \theta)\pi(p, \theta \mid \mathcal{Z}_n)dpd\theta.$$

On the basis of this formula, for each $(p^{(i)}, \theta^{(i)})$, $i = 1, \ldots, r$, one can simulate s processes until the lth generation, which started with Z_n individuals, obtaining the values $z^{(i)}_{n+l,1}$, $z^{(i)}_{n+l,2}$, …, $z^{(i)}_{n+l,s}$ from $Z_{n+l} > 0$, and use them to approximate $f(Z_{n+l} \mid \mathcal{Z}_n, p^{(i)}, \theta^{(i)})$ by a Gaussian kernel estimator

$$f^{(i)}(x) = \frac{1}{s}\sum_{j=1}^{s} \frac{1}{b^{(i)}} K\left(\frac{x - z^{(i)}_{n+l,j}}{b^{(i)}}\right), \; x \in \mathbb{R},$$

with $b^{(i)}$ an appropriate bandwidth and $K(x)$ the density of the standard normal distribution. Thus, $f(Z_{n+l} \mid \mathcal{Z}_n)$ is estimated by

$$\hat{f}(x) = \frac{1}{r}\sum_{i=1}^{r} f^{(i)}(x), \; x \in \mathbb{R}. \tag{11.10}$$

Method B We consider a random sample $\{(p^{(1)}, \theta^{(1)}), \ldots, (p^{(r)}, \theta^{(r)})\}$ (again the sample obtained in Sect. 11.2.2) from $\pi(p, \theta \mid \mathcal{Z}_n)$, and for each $i = 1, \ldots, r$ one simulates s processes until the lth generation, which started with Z_n, reproduction law $p^{(i)}$, and control distribution governed by $\theta^{(i)}$. One calculates the mean and the variance of the $s-$values $z^{(i)}_{n+l,j} > 0, j = 1, \ldots, s$, obtaining an approximation to $E[Z_{n+l} \mid Z_n, p^{(i)}, \theta^{(i)}]$ and to $Var[Z_{n+l} \mid Z_n, p^{(i)}, \theta^{(i)}]$. Finally,

$$E[Z_{n+l} \mid \mathcal{Z}_n] \approx \frac{1}{r}\sum_{i=1}^{r} E[Z_{n+l} \mid Z_n, p^{(i)}, \theta^{(i)}],$$

and, considering that

$$Var[Z_{n+l} \mid \mathcal{Z}_n] = E[Var[Z_{n+l} \mid \mathcal{Z}_n, p, \theta]] + Var[E[Z_{n+l} \mid \mathcal{Z}_n, p, \theta]],$$

with the mean and the variance in the right term of the previous equality considered with respect to the distribution $\pi(p, \theta \mid \mathcal{Z}_n)$, then

$$Var[Z_{n+l} \mid \mathcal{Z}_n] \approx \frac{1}{r}\sum_{i=1}^{r} Var[Z_{n+l} \mid Z_n, p^{(i)}, \theta^{(i)}] +$$

$$+\frac{1}{r-1}\sum_{i=1}^{r}\left(E[Z_{n+l} \mid Z_n, p^{(i)}, \theta^{(i)}] - \frac{1}{r}\sum_{i=1}^{r} E[Z_{n+l} \mid Z_n, p^{(i)}, \theta^{(i)}]\right)^2 .$$

As was already proposed in Mendoza and Gutiérrez-Peña [14] for BGW processes, we also use a gamma distribution with mean and variance $E[Z_{n+l} \mid \mathcal{Z}_n]$ and $Var[Z_{n+l} \mid \mathcal{Z}_n]$, respectively, (justified by the minimum logarithmic divergence criterion) to approximate $f(Z_{n+l} \mid \mathcal{Z}_n)$.

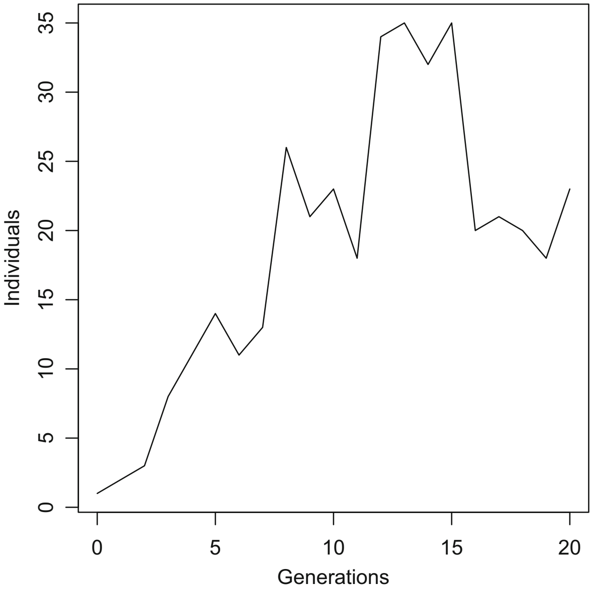
11.3 Simulated Example

In this section, we shall illustrate the methods described above by analysing an example with simulated data.

We simulated 20 generations of a CBP with $Z_0 = 1$, offspring law $p_0 = 0.0778$, $p_1 = 0.2592$, $p_2 = 0.3456$, $p_3 = 0.2304$, $p_4 = 0.0768$, $p_5 = 0.0102$, and $\phi_n(k)$ having a Poisson distribution with parameter $0.51k$. Thus the offspring mean is $m = 2$, the control and the migration parameters coincide, being $\mu = \theta = 0.51$, and in this case the asymptotic mean growth rate is $\tau = m\mu = 1.02$. As $\mu < 1$, we are considering a CBP with expected emigration. Figure 11.1 shows the evolution of the simulated population sizes. In an emulation of the classification of standard BGW processes, it was established in [8] that one refers to a subcritical, critical, or supercritical CBP depending on whether τ is less than, equal to, or greater than unity. Despite the expected emigration, using the results in [5], one can deduce that this supercritical CBP process has a positive probability of non-extinction.

We now focus on the estimation of p, μ, m, and τ based on the population size in each generation, by using the Gibbs sampler. We specify a Dirichlet process on the non-negative integers to model the prior distribution of the offspring law, avoiding any assumption about the cardinality of its support. Initially we choose $\alpha = 1$ as concentration parameter, and a Poisson distribution as base measure. It is well known that the Poisson distribution models the number of events occurring within a given time interval when those events occur at a known average rate and independently of the time since the last event. It is thus appropriate for modeling a generic offspring process. We propose that the average rate of the Poisson distribution will be initially estimated by considering that no control is imposed on the population. Therefore, the maximum likelihood estimator of the offspring mean corresponding to a BGW process can be used, i.e., we propose the Poisson base distribution with mean $(Z_1 + \ldots + Z_n)/(Z_0 + \ldots + Z_{n-1})$, in this example,

Fig. 11.1 Evolution of the simulated population sizes



1.06. Recall that for a Poisson control distribution—see Remark 11.1, b)—$\mu = \theta$, so that we shall henceforth refer in the analysis to μ. With respect to the prior for the parameter μ, as one does not know a priori what kind of expected migration is taking place, one can take the value of μ to be 1, and set as a prior distribution for μ a gamma distribution with mean 1 (the shape parameter is chosen to be unity). How the choices of prior elicitation affect the inferences will be evaluated below.

In the simulation, we set $T = 50$ and ran the algorithm described in Sect. 11.2.2 20,000 times for each chain. Using the Gelman–Rubin–Brooks diagnostic plots for (p, μ), we took $N = 5000$. Table 11.1 lists the values of the estimated potential scale reduction factor together with 97.5 % confidence upper bounds and the autocorrelation values for μ and the first values of p. That the values of the estimated scale reduction factor are close to unity suggests that further simulations will not improve the values of the listed scalar estimators (see [3, 4]). Finally, for the autocorrelation study, we chose $G = 600$, and consequently $Q = 25$. The final sample size was therefore 1300.

To evaluate the algorithm's efficiency, Table 11.2 presents some summary statistics for the posterior distributions of μ, m, and τ. Note that, due to the batch procedure, the time-series standard errors (TSSE) are very close to the Monte Carlo standard errors (MCSE). Also, for the three parameters, the standard errors (MCSE and TSSE) are less than 5 % of the posterior standard deviation (SD), indicating that the number of observations considered seems to be a reasonable choice.

Figures 11.2 and 11.3 show the estimated posterior density for μ, m, and τ together with their true values, and the 95 % HPD sets. The contour plot of the estimated posterior density for (μ, m) is also shown. One observes that the 95 %

Table 11.1 Potential scale reduction factor and autocorrelation for μ and the first values of p

	Potential scale reduction		Autocorrelation		
	Est.	97.5 %	lag1	lag100	lag 600
μ	1.04	1.06	0.9894	0.6669	0.0577
p_0	1.03	1.04	0.9937	0.6558	0.0776
p_1	1.02	1.03	0.9925	0.5538	0.0446
p_2	1.02	1.02	0.9895	0.4922	0.0326
p_3	1.02	1.04	0.9893	0.5055	0.0302
p_4	1.02	1.02	0.9684	0.1478	0.0020
p_5	1.03	1.04	0.9337	0.0500	0.0000
p_6	1.06	1.06	0.7524	0.0059	−0.0016

Table 11.2 Summary statistics for the posterior distributions of μ, m, and τ

	MEAN	SD	MCSE	TSSE
μ	1.0187	0.6015	0.0167	0.0155
m	1.2751	0.4987	0.0138	0.0127
τ	1.0597	0.0728	0.0020	0.0018

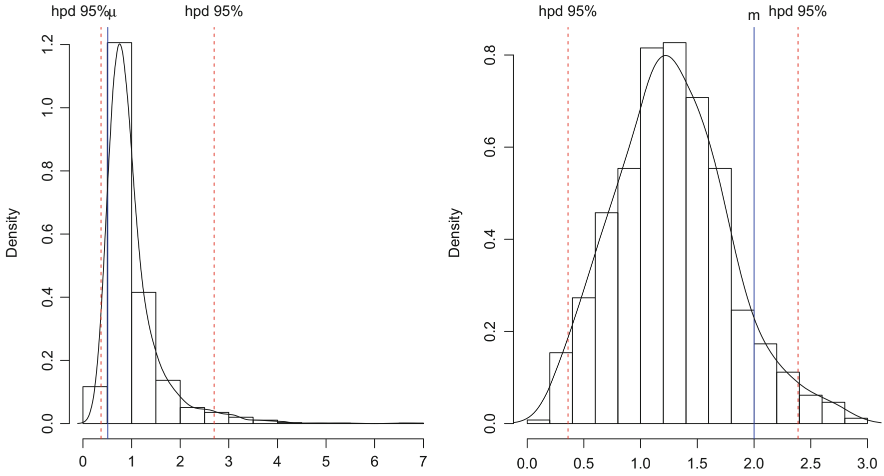


Fig. 11.2 Estimated posterior density for μ (*left*) and m (*right*)

HPD sets contain the true values of the parameters. It is particularly noteworthy that the method is not very accurate at identifying the parameters m and μ, although $P(\mu < 1 \mid \mathcal{Z}_{20})$ is estimated at 0.662, identifying the process as having expected emigration, and $P(m > 1 \mid \mathcal{Z}_{20})$ is estimated at 0.711, identifying a mean reproduction capacity of greater than unity. The method also provides a good estimate of the process's asymptotic mean growth rate, τ, which is the parameter that determines the limiting behaviour of the process. In this case, $P(\tau > 1 \mid \mathcal{Z}_{20})$ is approximated by 0.80, identifying a supercritical CBP. The contour plot shows clearly the interdependence of the parameters m, μ, and τ.

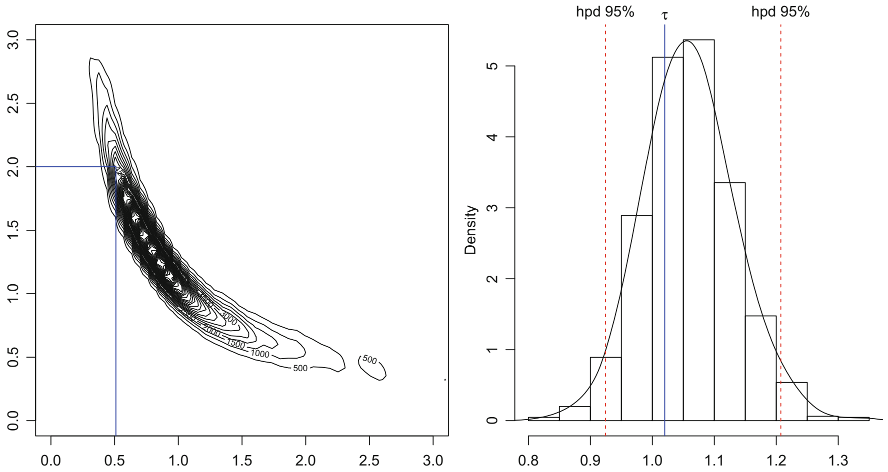


Fig. 11.3 Contour plot of estimated posterior density for (μ, m) (*left*) and estimated posterior density for τ (*right*)

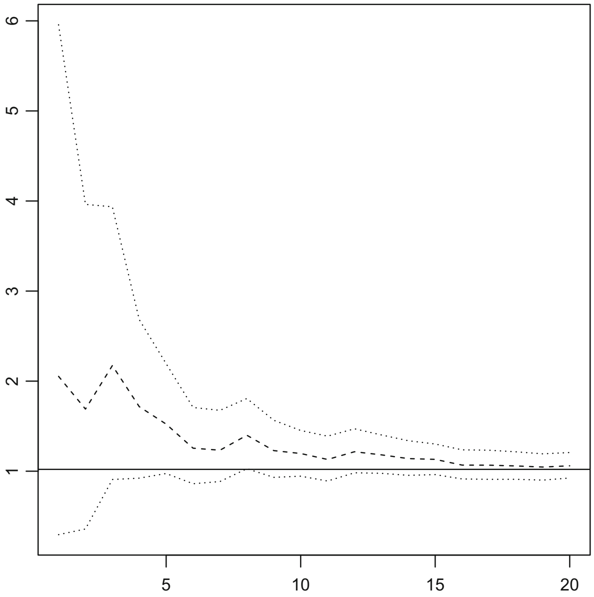


Fig. 11.4 Evolution of the squared error loss estimates of τ with 95 % HPD bands. The *horizontal line* represents the true parameter value

Figure 11.4 illustrates the *long-term* behaviour of the estimates of τ, showing for each generation their Bayes estimates under squared error loss and their respective 95 % HPD sets. Note that one has estimates closer to the real value and narrower HPD intervals as the generations advance.

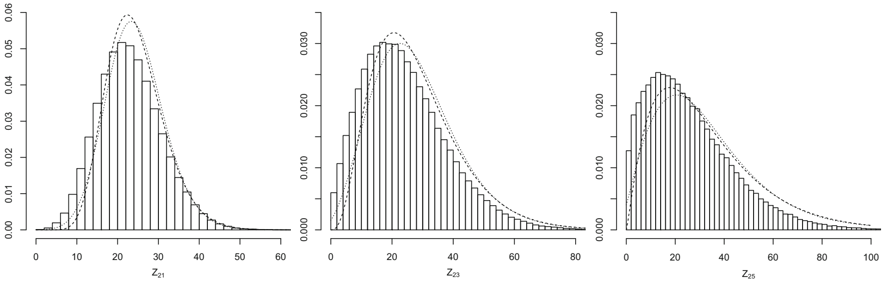


Fig. 11.5 Histogram of the data from the true distribution together with estimated posterior densities for Z_{21}, Z_{23}, and Z_{25}. Method A: *dotted line*; Method B: *discontinuous line*

Table 11.3 Mean and 95 % prediction HPD sets for Z_{21}, Z_{23}, and Z_{25}

	MEAN Method A/Method B	95 % prediction HPD Method A/Method B
Z_{21}	24.3844 / 24.3840	11.2112–38.2382 /11.6116–38.7387
Z_{23}	27.8852 / 27.8475	2.8028–57.8579 /4.6046–58.9590
Z_{25}	31.7907 / 32.6389	3.9039–86.9870/1.1011–84.4845

Using the information on the population sizes until generation 20, we obtained via the methods described in Sect. 11.2.3 the predictive distributions of Z_{21}, Z_{23}, and Z_{25} on their respective non-extinction sets. Methods A and B were applied by simulating $s = 1000$ processes, started with $Z_{20} = 23$ until the 5th generation, and reproduction law and control parameter $(p^{(i)}, \theta^{(i)})$, $i = 1, \ldots, 1300$.

Figure 11.5 shows the predictive posterior distribution of Z_{21}, Z_{23}, and Z_{25} estimated by the two methods described above, together with the histogram of the data from the true distribution. The sampling-based approach and the gamma model lead to similar estimates for Z_{21}, Z_{23}, and Z_{25}. Table 11.3 presents estimates for Z_{21}, Z_{23}, and Z_{25} together with the 95 % HPD sets. The accuracy of these approximations, assessed by comparison with the true distribution, is reasonable. In spite of the fact that the estimation of m and μ are not so precise as desirable, the combination of both to estimate τ is quite good, as shown Fig. 11.3 (right) and this leads to the estimation of the forecast values of the process are accurate. Nevertheless, according to Fig. 11.3 (right), the estimated values of τ tend to be slightly greater than the true value, so that the predictive values also show this tendency.

Finally, we examine the sensitivity of inferences of the main parameters of interest to the choice of the priors (in particular, focusing on the concentration parameter, the base distribution, and the gamma parameters). For simplicity, we present the analysis focussing on the asymptotic mean growth rate, τ, a parameter that determines the future evolution of the process. First, we analyse the influence on the choices of the concentration parameter and of the kind of base distribution (setting its mean to 1.06). With the prior on μ assumed to be a gamma distribution

Table 11.4 Sensitivity analysis for $\tau|\mathcal{Z}_{20}$

Concentration parameter α	Base measure $m^{(0)} = 1.06$					
	Poisson			Geometric		
	MEAN	HPD 95 %		MEAN	HPD 95 %	
0.25	1.0629	0.9244	1.2060	1.0584	0.9164	1.2107
0.50	1.0646	0.9258	1.2136	1.0626	0.9251	1.2122
0.75	1.0616	0.9195	1.2118	1.0618	0.9187	1.2180
1	1.0597	0.9244	1.2073	1.0585	0.9053	1.2129
5	1.0583	0.9249	1.2104	1.0618	0.9062	1.2271
10	1.0613	0.9204	1.2081	1.0610	0.8980	1.2194
20	1.0554	0.9099	1.2041	1.0642	0.9065	1.2391

Table 11.5 Sensitivity analysis for $\tau|\mathcal{Z}_{20}$

$m^{(0)}$	Base measure $\alpha = 1$					
	Poisson			Geometric		
	MEAN	HPD 95 %		MEAN	HPD 95 %	
0.25	1.05592	0.93007	1.19095	1.05441	0.92359	1.19157
0.50	1.05687	0.92847	1.19287	1.05850	0.92709	1.19397
0.75	1.05983	0.93648	1.19943	1.06178	0.91882	1.21939
1	1.05664	0.91961	1.21131	1.05850	0.91213	1.20456
1.5	1.06283	0.91825	1.22327	1.06127	0.90646	1.22152
2	1.06270	0.90100	1.21972	1.06129	0.91375	1.21722
2.5	1.06631	0.91151	1.22647	1.06287	0.90328	1.24079
3	1.06887	0.90780	1.23867	1.06043	0.90816	1.23099
3.5	1.06978	0.89095	1.24436	1.06787	0.90444	1.25410

with mean 1 (its shape parameter taken as unity), the results in Table 11.4 show the estimation of τ not to be very sensitive to such changes. Second, with the same prior on μ and the concentration parameter equal to unity, the results in Table 11.5 show the estimation of τ neither to be very sensitive to the choice of the base distribution or of its mean. Finally, we analyse the influence of the choice of the shape and mean parameters of the gamma distribution. Taking the previous study into account, we took the concentration parameter to be equal to unity, and the base distribution to be a Poisson distribution with mean 1.06. The results in Table 11.6 again allow one to conclude that the estimation of τ is not very sensitive to the prior parameters.

In most of the situations, as was noted in Remark 11.3, one has prior knowledge of the kind of the control distribution, and this was indeed our approach in the simulated example where we assumed the control to be applied through a Poisson control distribution. We next examined how the method works when assuming complete ignorance of the control law. To this end, we implemented the method by considering prior binomial, negative binomial, and Poisson control laws. Thus, besides the data already simulated and presented in Fig. 11.1 which corresponded to an expected emigration, we also considered a new sample from a CBP with Poisson

Table 11.6 Sensitivity analysis for $\tau|\mathcal{Z}_{20}$

Gamma distribution								
Mean parameter		Shape parameter						
		0.25	0.50	0.75	1	2.5	5	7.5
0.25	MEAN	1.0607	1.0546	1.0563	1.0506	1.0408	1.0293	1.0207
	95 % HPD lower bound	0.9234	0.9021	0.9138	0.9023	0.8980	0.8841	0.8665
	95 % HPD upper bound	1.2075	1.2219	1.1988	1.2153	1.1848	1.1860	1.1937
0.5	MEAN	1.0597	1.0524	1.0570	1.0523	1.0518	1.0466	1.0453
	95% HPD lower bound	0.9168	0.9189	0.9052	0.9152	0.9047	0.9101	0.8998
	95 % HPD upper bound	1.2128	1.1997	1.2160	1.2063	1.2066	1.1950	1.1911
0.75	MEAN	1.0638	1.0660	1.0632	1.0598	1.0571	1.0596	1.0575
	95 % HPD lower bound	0.9242	0.9242	0.9198	0.9233	0.9177	0.9055	0.9194
	95 % HPD upper bound	1.2065	1.2315	1.2136	1.2043	1.2042	1.2133	1.2075
1	MEAN	1.0617	1.0613	1.0606	1.0597	1.0620	1.0629	1.0689
	95 % HPD lower bound	0.9193	0.9146	0.9164	0.9244	0.9191	0.9219	0.9268
	95 % HPD upper bound	1.2132	1.2144	1.2119	1.2073	1.2121	1.2125	1.2058
5	MEAN	1.0599	1.0609	1.0652	1.0662	1.0628	1.0596	1.0657
	95 % HPD lower bound	0.9042	0.9247	0.9302	0.9289	0.9241	0.9185	0.9250
	95 % HPD upper bound	1.2102	1.2045	1.2124	1.2083	1.2107	1.2102	1.2145
10	MEAN	1.0617	1.0660	1.0620	1.0641	1.0607	1.0659	1.0634
	95 % HPD lower bound	0.9179	0.9225	0.9188	0.9273	0.9212	0.9319	0.9240
	95 % HPD upper bound	1.2101	1.2128	1.2089	1.2037	1.1952	1.2101	1.2073
20	MEAN	1.0590	1.0630	1.0644	1.0603	1.0632	1.0633	1.0662
	95 % HPD lower bound	0.9232	0.9323	0.9195	0.9131	0.9259	0.9355	0.9354
	95 % HPD upper bound	1.2109	1.2084	1.2042	1.2082	1.2099	1.1960	1.2092

control distributions with the same asymptotic mean growth rate as the previous example but now with an expected immigration (with parameters $m = 0.51$ and $\mu = 2$). The results in Table 11.7 illustrate how well (or now poorly) the fitted models, in both the expected emigration and the expected immigration examples, identify the asymptotic mean growth rate of the process as well as a supercritical or a subcritical reproduction mean (i.e., $m > 1$ or $m < 1$, respectively). To assess whether the fitted models detect an expected emigration or immigration, we calculated $P(\mu < 1 \mid \mathcal{Z}_{20})$ or $P(\mu > 1 \mid \mathcal{Z}_{20})$, respectively. The approximations of these probabilities are also given in the table. The results in the table show the asymptotic mean growth rate of the process to usually be well identified whichever situation is considered. This is because the generation-by-generation population sizes provide enough information to estimate this value. However, the estimates of the kind of offspring mean and of the migration process are not generally appropriate in the absence of prior knowledge about the type of control. Analysing the fitted model considering a prior Poisson control (the genuine control distribution), we obtained acceptable results for the estimates of the offspring mean and of the expected migration process in both the emigration and the immigration examples.

Table 11.7 Sensitivity analysis: control prior distribution

Simulated model	Fitted models					
Poisson control	P control		B control		NB control	
Expected emigration	$\tau \mid \mathcal{Z}_{20}$					
$m = 2;\ \tau = 1.02$	Mean	Variance	Mean	Variance	Mean	Variance
	1.060	0.005	1.054	0.005	1.062	0.007
	$P(m > 1 \mid \mathcal{Z}_{20})$					
	0.711		0.995		0.400	
	$P(\mu < 1 \mid \mathcal{Z}_{20})$					
	0.662				0.317	
Expected immigration	$\tau \mid \mathcal{Z}_{20}$					
$m = 0.51;\ \tau = 1.02$	Mean	Variance	Mean	Variance	Mean	Variance
	1.061	0.006	1.056	0.007	1.064	0.012
	$P(m < 1 \mid \mathcal{Z}_{20})$					
	0.626		0.013		0.735	
	$P(\mu > 1 \mid \mathcal{Z}_{20})$					
	0.767				0.801	

P control Poisson control, *B control* binomial control, *NB control* negative binomial control

Considering a prior binomial control, we found a tendency to overestimate the offspring mean for the two simulated data sets. This was especially so in case of the expected immigration situation, precisely to compensate for the effect of the immigration (recall that the binomial control only allows for emigration). Finally, with a negative binomial control, one deduces from the results in Table 11.7 that, despite both the simulated and the fitted models allowing for any kind of migration, the fitted model identifies well neither the offspring mean nor the emigration process when the latter is expected. The conclusion to be drawn from these simulated examples is thus that, to obtain reliable estimates of the parameters of interest, one should have some prior knowledge of the kind of control law.

Remark 11.4 For the computation of the examples, we used the statistical software **R**, a language and environment for statistical computing and graphics (see [16]), performing the simulations by parallel computing using the Rmpi [25] and snow (see [21]) packages, and the convergence diagnostics using the coda package (see [15]).

11.4 Concluding Remarks

As was noted in the Introduction, nowadays controlled branching processes are one of the most relevant branching models for study. Apart from their inherent interest, this current relevance is due to the fact that they generalize many important branching processes—migration models, for instance. Moreover, beyond the framework of population dynamics, these models cover and generalize popular models for the analysis of count-data time series, such as the INAR(1) models (see [22, 23] for reviews). Consequently, the development of the inferential theory of CBPs is an important task to provide a guarantee for their application to many real-world problems. Some contributions to this theory have already been made from a frequentist standpoint, either for the general model itself ([6, 19]) or for some of its particular cases, such as the models with immigration (see, e.g., [17] and the references therein) or INAR(1) (see [24]).

Respect to a Bayesian perspective, this chapter extends in several senses the preliminary contributions established in [9, 10]. In particular, we have focused on a CBP with non-parametric offspring law and with random control variables instead of deterministic ones, assuming these belonging to a parametric family of distributions that depend on a single parameter, termed the control parameter (or its equivalent, the migration parameter). To avoid any assumption about the cardinality of the support of the offspring law, a Dirichlet process was introduced. The classical application of this methodological approach to a branching context in a non-parametric framework for the offspring law requires the observation of the entire family tree (up to some generation). Although we developed this approach -with the novelty of using the Dirichlet process-, we reckon that from a practical standpoint it is more realistic and relevant to avoid the need for observation of the complete family tree and we only consider the record of the total generation sizes. On the basis of this kind of sample, and making use of the Gibbs sampler and kernel density estimators, we proposed a method to approximate the posterior density of the control parameter (or of its parametrization as the migration parameter), of the offspring law, and consequently the posterior densities of the offspring mean and asymptotic mean growth rate, regardless of the nature of the offspring law (whether subcritical, critical, or supercritical) and the kind of migration. It is worth noting that the frequentist methods given until now depend strongly on this nature, making a Bayesian method preferable in this sense. In fact, one can make inferences on the nature of the offspring law or the kind of migration (immigration or emigration) from the posterior density of the offspring mean or of the migration parameter, respectively.

Furthermore, the above method also allows one to approximate the predictive posterior densities. Two methods were proposed to make inferences on the size of future generations.

We considered a simulated example in some depth to illustrate our findings, and included a detailed sensitivity analysis regarding the choices of the priors. This showed the methodological approach to not be unduly influenced by the choice of priors of the control parameter or the priors of the concentration parameter or the base distribution. However, it revealed the need for prior knowledge of the kind of control being applied. With respect to the comparison of the two methods for predictions, the simulated example showed the two approaches to lead essentially to the same results.

As an overall conclusion, we would state that the proposed procedure allowing inference based only on total generation sizes constitutes the main contribution of the present work.

Acknowledgements The authors would like to thank Professors Horacio González-Velasco and Carlos García-Orellana for providing them with computational support. This research has been supported by the Ministerio de Educación, Cultura y Deporte (grant FPU13/03213), Ministerio de Economía y Competitividad of Spain (grants MTM2012-31235 and MTM2015-70522-P), the Junta de Extremadura (grant GR15105), the FEDER, and the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria (grant DFNI-I02/17).

References

1. Bagley, J.H.: On the almost sure convergence of controlled branching processes. J. Appl. Probab. **23**, 827–831 (1986)
2. Brooks, S.: Markov Chain Monte Carlo method and its application. J. R. Stat. Soc. Ser. D (The Statistician) **47**, 69–100 (1998)
3. Brooks, S., Gelman, A.: General methods for monitoring convergence of iterative simulations. J. Comput. Graph. Stat. **7**, 434–455 (1998)
4. Gelman, A., Rubin, D.: Inference from iterative simulation using multiple sequences. Stat. Sci. **7**, 457–511 (1992)
5. González, M., Molina, M., del Puerto, I.: On the class of controlled branching process with random control functions. J. Appl. Probab. **39**, 804–815 (2002)
6. González, M., Martínez, R., del Puerto, I.: Nonparametric estimation of the offspring distribution and mean for a controlled branching process. Test **13**, 465–479 (2004)
7. González, M., Martínez, R., del Puerto, I.: Estimation of the variance for a controlled branching process. Test **14**, 199–213 (2005)
8. González, M., Molina, M., del Puerto, I.: Asymptotic behaviour of critical controlled branching process with random control function. J. Appl. Probab. **42**(2), 463–477 (2005)
9. González, M., Martín, J., Martínez, R., Mota, M.: Non-parametric Bayesian estimation for multitype branching processes through simulation-based methods. Comput. Stat. Data Anal. **52**, 1281–1291 (2008)
10. González, M., Gutiérrez, C., Martínez, R., del Puerto, I.: Bayesian inference for controlled branching processes through MCMC and ABC methodologies. Rev. R. Acad. Cien. Serie A. Mat. **107**, 459–473 (2013)
11. González, M., Minuesa, C., del Puerto, I.: Maximum likelihood estimation and Expectation-Maximization algorithm for controlled branching process. Comput. Stat. Data Anal. **93**, 209–227 (2016)
12. Jagers, P.: Branching Processes with Biological Applications. Wiley, London (1975)
13. Martínez, R., Mota, M., del Puerto, I.: On asymptotic posterior normality for controlled branching processes with random control function. Statistics **43**, 367–378 (2009)
14. Mendoza, M., Gutiérrez-Peña, E.: Bayesian conjugate analysis of the Galton-Watson process. Test **9**, 149–171 (2000)
15. Plummer, M., Best, N., Cowles, K., Vines, K.: coda: Output analysis and diagnostics for MCMC. R package version 0.13-5 (2010)
16. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2014)
17. Rahimov, I.: Limit distributions for weighted estimators of the offspring mean in a branching process. Test **18**, 568–583 (2009)

18. Sevastyanov, B.A., Zubkov, A.: Controlled branching processes. Theor. Prob. Appl. **19**, 14–24 (1974)
19. Sriram, T., Bhattacharya, A., González, M., Martínez, R., del Puerto, I.: Estimation of the offspring mean in a controlled branching process with a random control function. Stoch. Process. Appl. **117**, 928–946 (2007)
20. Tierney, L.: Markov chains for exploring posterior distributions. Ann. Stat. **22**, 1701–1762 (1994)
21. Tierney, L., Rossini, A.J., Li, N., Sevcikova, H.: snow: Simple Network of Workstations. R package version 0.3-3 (2013)
22. Tjøstheim, D.: Some recent theory for autoregressive count time series. Test **21**, 413–438 (2012)
23. Weiß, C.H.: Thinning operations for modelling time series of counts- a survey. Adv. Stat. Anal. **92**, 319–341 (2008)
24. Weiß, C.H.: Fully observed INAR(1) processes. J. Appl. Stat. **39**, 581–598 (2011)
25. Yu, H.: Rmpi: parallel statistical computing in R. R News **2**, 10–14 (2002)

Chapter 12
Recurrence and Transience of Near-Critical Multivariate Growth Models: Criteria and Examples

Götz Kersting

Mathematics Subject Classification (2000): 60J10, 60J80

12.1 Introduction

We consider stochastic processes $(X_n)_{n\geq 0}$ taking values in $\mathbb{R}_+^d = \{(x_1,\ldots,x_d)^T \in \mathbb{R}^d : x_i \geq 0\}$, i.e. the d-dimensional orthant, and adapted to some filtration $(\mathcal{F}_n)_{n\geq 0}$, which satisfy an equation of the form

$$X_{n+1} = MX_n + g(X_n) + \xi_n \ , \quad n \in \mathbb{N}_0 \ ,$$

with a $d \times d$ matrix M having non-negative entries, with a function $g : \mathbb{R}_+^d \to \mathbb{R}^d$, and with random fluctuations $\xi_n = (\xi_{n1},\ldots,\xi_{nd})^T$ satisfying

$$\mathbf{E}[\xi_n \mid \mathcal{F}_n] = 0 \ \text{ a.s.}$$

One may view the process as a non-linear random perturbation of the linear dynamical system $x_{n+1} = Mx_n$, $n \in \mathbb{N}_0$. Here we require this system to be critical, which means that the Perron-Frobenius eigenvalue λ_1 of M is equal to 1. We focus on the situation, when M is a primitive matrix; then up to scaling there is a unique left eigenvector $\ell = (\ell_1,\ldots,\ell_d)$ corresponding to λ_1 and its components are all strictly positive. As usual we let $\ell_1 + \cdots + \ell_d = 1$.

G. Kersting (✉)
Institut für Mathematik, Goethe Universität, Frankfurt am Main, Frankfurt, Germany
e-mail: kersting@math.uni-frankfurt.de

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_12

Now the size of the random fluctuations will be determined by the conditional variance of $\ell\xi_n = \ell_1\xi_{n1} + \cdots + \ell_d\xi_{nd}$. More precisely we assume that

$$\mathbf{E}[(\ell\xi_n)^2 \mid \mathcal{F}_n] = \sigma^2(X_n) \ \text{ a.s.}$$

with some measurable function $\sigma^2 : \mathbb{R}_+^d \to \mathbb{R}_+$.

We are aiming at general criteria for recurrence or transience of the process $(X_n)_{n\geq 0}$, that is, at criteria which allow to decide whether $\{\|X_n\| \to \infty \text{ for } n \to \infty\}$ is an event of zero probability or not, where $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^d$. Hereby, in talking about recurrence and transience, we have taken the liberty to adopt the terminology from Markov chain theory. Certainly our processes obey only a relaxed form of the Markov property, but examples typically are Markov chains.

Our theorems in the multivariate setting are complete generalizations of the known results in the univariate setting. Therefore it is appropriate to first reconsider the univariate case. This is done in Sect. 12.2. In Sect. 12.3 we apply these results to the population size dependent bisexual Galton-Watson process. The multivariate case is then discussed in Sect. 12.4. As an example the multivariate population size dependent Galton Watson process is treated in Sect. 12.5. Proofs for the multivariate criteria are given in Kersting (2016), arxiv:1605.04064 [math.PR].

12.2 The Univariate Case Revisited

In the 1-dimensional case our model equation simplifies to the difference equation

$$X_{n+1} = X_n + g(X_n) + \xi_n$$

with some function $g : \mathbb{R}_+ \to \mathbb{R}$. A number of examples can be put into this framework, among others e.g. population size dependent branching processes [10, 12], controlled branching processes [5], branching processes in random environment [2], or nonlinear stochastic trends [7].

Our main condition is

$$g^+(x) = o(x) \ \text{ as } x \to \infty \ . \tag{A1}$$

This assumption of "near-criticality" simply says that X_n is the dominating term within $X_n + g(X_n)$ such that supercritical growth is excluded.

Also we assume the existence of some constants $c, \delta > 0$ such that for $X_n \geq c$

$$\mathbf{E}[|\xi_n|^p \mid \mathcal{F}_n] \leq c\sigma^p(X_n) \ \text{ a.s. with } p = 2 + \delta \ . \tag{A2}$$

With these two conditions we have the following criteria complementary to each other.

Theorem 12.1 *Let* (A1), (A2) *be fulfilled. Assume that there is an* $\varepsilon > 0$ *such that*

$$xg(x) \leq \frac{1-\varepsilon}{2}\sigma^2(x) \tag{12.1}$$

for x *sufficiently large. Then*

$$\mathbf{P}(X_n \to \infty \, \textit{for} \, n \to \infty) = 0 \, .$$

The converse criterium requires some slight additional restrictions.

Theorem 12.2 *Let* (A1), (A2) *be fulfilled. Also let* $\sigma^2(x)$ *be bounded away from zero on intervals* (u, v) *with* $0 < u < v < \infty$*, and let*

$$\sigma^2(x) = O(x^2 \log^{-2/\delta} x) \;\; \textit{for}\, x \to \infty \tag{12.2}$$

with δ *as in* (A2). *Assume that there is an* $\varepsilon > 0$ *such that*

$$xg(x) \geq \frac{1+\varepsilon}{2}\sigma^2(x) \tag{12.3}$$

for x *sufficiently large. Then there is a number* $m < \infty$ *such that*

$$\mathbf{P}(\limsup_n X_n \leq m \; \textit{or} \; \lim_n X_n = \infty) = 1 \, .$$

If also for every $c > 0$ *there is a* $n \in \mathbb{N}$ *such that* $\mathbf{P}(X_n > c) > 0$*, then*

$$\mathbf{P}(X_n \to \infty \, \textit{for} \, n \to \infty) > 0 \, .$$

These results are contained in [9]. From there the first theorem is taken literally, while the second one is a somewhat more general version of the corresponding Theorem 2 in [9]. There the condition $g(x) = O(x \log^{-2/\delta} x)$ is used, which is stronger than our condition (12.2) in view of (12.3). Thus our theorem offers a relaxation of conditions, which is more to the point and also useful in examples, while the proof of the criterion remains practically unchanged (as one easily convinces oneself).

The theorems can be understood as follows: Typically the long term behavior is either dominated by the "drift term" $g(X_n)$, or it is mainly controlled by the fluctuations ξ_n. There is only a small boundary region where both the drift term and the fluctuations have to be taken into account. It is there, where one would expect a particular rich and variable stochastic behavior.

Remark 12.1 In order to get a better understanding of the main condition of both theorems it is instructive to rewrite the model equation in a multiplicative form as

$$X_{n+1} = X_n(1 + h(X_n) + \zeta_n) \quad \text{with } h(x) = \frac{g(x)}{x} \, , \; \zeta_n = \frac{\xi_n}{X_n} \, .$$

Then

$$\mathbf{E}[\zeta_n^2 \mid \mathcal{F}_n] = \tau^2(X_n) \text{ a.s.} \quad \text{with } \tau^2(x) = \frac{\sigma^2(x)}{x^2} .$$

Now the main requirements (12.1) and (12.3) of the theorems read $h(x) \leq \frac{1-\varepsilon}{2}\tau^2(x)$ versus $h(x) \geq \frac{1+\varepsilon}{2}\tau^2(x)$. In this formulation the drift is directly related to the variance of the fluctuations.

Remark 12.2 One might wonder, whether condition (12.2) can be substantially relaxed or even removed. Without compensation this is not possible, as can be seen from Example C in Sect. 3 of [9].

12.3 Example: The Bisexual GW-Process

For the bisexual Galton-Watson process the n-th generation of some population (with $n = 0, 1, \ldots$) consists of F_n female and M_n male individuals. They are assumed to form $L(F_n, M_n)$ different couples, with some given deterministic "mating function" $L(\cdot, \cdot)$, such that $L(0, \cdot) = L(\cdot, 0) = 0$. The i-th couple then has ρ_{ni} female and τ_{ni} male offspring. Thus the population evolves according to the equations

$$F_{n+1} = \sum_{i=1}^{L(F_n, M_n)} \rho_{ni} , \quad M_{n+1} = \sum_{i=1}^{L(F_n, M_n)} \tau_{ni} .$$

Let $\mathcal{F}_n$ be the σ-field, generated by the random pairs (F_k, M_k), $k = 0, \ldots, n$. For every $n \geq 0$ we assume that, given $\mathcal{F}_n$, the pairs (ρ_{ni}, τ_{ni}), $i \geq 1$, are iid random variables with values in $\mathbb{N}_0 \times \mathbb{N}_0$. In former investigations it has been assumed that their conditional distribution μ_n is non-random, here we allow that μ_n depends on $L(F_n, M_n)$. Thus we deal with a population size dependent bisexual Galton-Watson process. Note that the random variables $X_n = L(F_n, M_n)$, $n = 0, 1, \ldots$, form a Markov chain with values in $\mathbb{N}_0$ and an absorbing state 0.

For the bisexual Galton-Watson without population size dependence the question of recurrence/transience (or in other words the question, whether extinction appears with probability 1) has been completely solved for the large class of superadditive mating functions, see [3] and the literature cited therein. In [13, 15] the authors treated the case of mating functions depending on the population size.

Here we consider the situation where the distribution of (ρ_{ni}, τ_{ni}) may depend on the number of couples X_n (this is a particular case of the model introduced in [14]). Then it is necessary to specify the function L in more detail. We consider the prominent case

$$L(x, y) = \min(x, ry),$$

where $r \geq 1$ is a natural number ($r = 1$ means monogamous mating and $r \geq 2$ polygamous mating). We restrict ourselves to the balanced situation when

$$\mathbf{E}[\rho_{ni} \mid \mathcal{F}_n] = r\,\mathbf{E}[\tau_{ni} \mid \mathcal{F}_n] \text{ a.s.}$$

(the unbalanced case can be treated equally). In the case $r = 1$ this means that in mean all females and males will find together in couples and it is only due to random fluctuations that some will not succeed.

We like to apply our theorems to the process $(X_n)_{n \geq 0}$. The function g evaluated on $x > 0$ is given by

$$x + g(x) = \mathbf{E}[X_{n+1} \mid X_n = x] = \mathbf{E}_x\Big[\min\Big(\sum_{i=1}^{x} \rho_{0i}, r\sum_{i=1}^{x} \tau_{0i}\Big)\Big] ,$$

where we now use the notation $\mathbf{E}_x[\,\cdot\,] = \mathbf{E}[\,\cdot \mid X_0 = x]$ familiar for Markov chains. Denoting

$$e(x) = \mathbf{E}_x[\rho_{01}]$$

and using the identity $2\min(u, v) = u + v - |u - v|$, we may rewrite the above equation as

$$g(x) + x = xe(x) - \tfrac{1}{2}\mathbf{E}_x\Big[\Big|\sum_{i=1}^{x}(\rho_{0i} - r\tau_{0i})\Big|\Big]$$

in the balanced case. The right-hand expectation can be asymptotically evaluated as follows: Assuming that the function

$$v(x) = \mathbf{E}_x[(\rho_{01} - r\tau_{01})^2]$$

has a finite, strictly positive limit for $x \to \infty$, i.e.

$$v(x) \to \alpha > 0 \quad \text{as } x \to \infty ,$$

and assuming also

$$\mathbf{E}_x[\rho_{01}^{2+\eta} + \tau_{01}^{2+\eta}] \leq c \tag{12.4}$$

for some $\eta > 0$, $c < \infty$, we deduce from Lyapunov's version of the central limit theorem that

$$\mathbf{E}_x\Big[\Big|\sum_{i=1}^{x}(\rho_{0i} - r\tau_{0i})\Big|\Big] = \sqrt{\alpha x}(\mathbf{E}[|N|] + o(1)) \quad \text{as } x \to \infty ,$$

where N has a standard normal distribution. Since the right-hand expectation is equal to $\sqrt{2/\pi}$, we end up with

$$g(x) = (e(x)-1)x - \sqrt{\frac{\alpha x}{2\pi}} + o(\sqrt{x}) \quad \text{as } x \to \infty\,.$$

The moments of ξ_n can be obtained as follows:

$$\begin{aligned}
&\mathbf{E}_x[|\xi_0|^{2+\delta}] \\
&\quad = \mathbf{E}_x\Big[\Big|\min\Big(\sum_{i=1}^{x}\rho_{0i}, r\sum_{i=1}^{x}\tau_{0i}\Big) - x - g(x)\Big|^{2+\delta}\Big] \\
&\quad = \mathbf{E}_x\Big[\Big|\min\Big(\sum_{i=1}^{x}(\rho_{0i}-e(x)), \sum_{i=1}^{x}(r\tau_{0i}-e(x))\Big) + \sqrt{\frac{\alpha x}{2\pi}} + o(\sqrt{x})\Big|^{2+\delta}\Big]
\end{aligned}$$

From the Marcinkiewicz-Zygmund and the Hölder inequality we have

$$\begin{aligned}
\mathbf{E}_x\Big[\Big|\sum_{i=1}^{x}(\rho_{0i}-e(x))\Big|^{2+\eta}\Big] &\le c\mathbf{E}\Big[\Big(\sum_{i=1}^{x}|\rho_{0i}-e(x)|^2\Big)^{1+\eta/2}\Big] \\
&\le c\mathbf{E}\Big[\sum_{i=1}^{x}|\rho_{0i}-e(x)|^{2+\eta}x^{\eta/2}\Big] = cx^{1+\eta/2}\mathbf{E}[|\rho_{01}-e(x)|^{2+\eta}]
\end{aligned}$$

for some $c > 0$. Similarly the moment of the other sum may be estimated from above. Thus because of (12.4) we may apply the multivariate central limit theorem to obtain for $0 \le \delta < \eta$

$$\mathbf{E}_x[|\xi_0|^{2+\delta}] = x^{1+\delta/2}\mathbf{E}\Big[\Big|\min(N_1,N_2) + \sqrt{\alpha/2\pi}\Big|^{2+\delta}\Big] + o(x^{1+\delta/2})$$

as $x \to \infty$, where the distribution of (N_1, N_2) is bivariate normal. In particular the right-hand expectation is strictly positive. Since we are dealing with a time homogeneous Markov chain, (A2) is fulfilled for all $\delta < \eta$.

We are now ready to apply our theorems. The above formulas suggest to chose

$$e(x) = 1 + \frac{\beta}{\sqrt{x}} + o(x^{-1/2}) \ \text{ as } x \to \infty$$

for some real number β. Then asymptotically

$$g(x) \sim (\beta - \sqrt{\alpha/2\pi})\sqrt{x}\,, \quad \sigma^2(x) = \mathbf{E}_x[\xi_0^2] \sim \gamma x$$

for some $\gamma > 0$. Thus applying Theorems 12.1 and 12.2 and noting that 0 is the only absorbing state, we end up with the following result.

Corollary 12.1 *Assume that*

$$e(x) = 1 + \frac{\beta}{\sqrt{x}} + o(x^{-1/2}) \ \ as\ x \to \infty$$

and that $\mathbf{E}_x[\rho_{01}^{2+\eta} + \tau_{01}^{2+\eta}] \leq c$ *for all* x *large enough and some* $\eta, c > 0$. *Then we have:*

(i) *If* $\beta < \sqrt{\alpha/2\pi}$, *then the process* $(X_n)_{n\geq 0}$ *gets extinct with probability 1.*
(ii) *If* $\beta > \sqrt{\alpha/2\pi}$, *then* $(X_n)_{n\geq 0}$ *diverges with positive probability.*

12.4 The General Case

Now we come back to the d-dimensional process $(X_n)_{n\geq 0}$ with values in $\mathbb{R}_+^d$ satisfying

$$X_{n+1} = MX_n + g(X_n) + \xi_n \ \text{ with } \mathbf{E}[\xi_n \mid \mathcal{F}_n] = 0\ ,\ \mathbf{E}[(\ell\xi_n)^2 \mid \mathcal{F}_n] = \sigma^2(X_n) \text{ a.s.}$$

We recall that M is assumed to be a primitive matrix, that is its entries are non-negative and there is a natural number k such that the entries of M^k are all strictly positive. We further assume that the Perron-Frobenius eigenvalue of M is equal to 1. As is well-known, see [16], it has unique left and right eigenvectors $\ell = (\ell_1, \dots, \ell_d)$ and $r = (r_1, \dots, r_d)^T$ with strictly positive components and normalized by $\ell_1 + \cdots + \ell_d = \ell_1 r_1 + \cdots + \ell_d r_d = 1$.

We like to obtain generalisations of the Theorems 12.1 and 12.2 above. One approach is to look for suitable d-dimensional versions of the recurrence condition (12.1) resp. of the transience condition (12.3) holding everywhere in the $\mathbb{R}_+^d$ (possibly up to a neighbourhood of the origin). This approach has been performed by González et al [4], who considered in fact more general multitype Markov chains and applied their results to controlled branching processes (see also [6]).

On the other hand there occur interesting instances where in some parts of the state space one then would come across the recurrence condition (12.1) and in others across the transience condition (12.3). Therefore we follow a different idea. From Perron-Frobenius theory it is known that the modulus of the other eigenvalues of M are all smaller than 1. Roughly speaking this implies that after multiplication with M vectors $x \in \mathbb{R}_+^d$ are pushes towards the direction of the eigenvector r. The same effect is also active for the process $(X_n)_{n\geq 0}$ (provided it is not nullified by the presence of g or the ξ_n). Thus on the event $\|X_n\| \to \infty$ one would expect that the sequence X_n diverges asymptotically in the direction determined by r. Consequently the mentioned recurrence or transience conditions have to be required only in some vicinity of the ray $\overline{r} = \{\lambda r : \lambda \geq 0\}$ spanned by the vector r. This is the type of condition we are aiming at. First results in this direction are due to Adam [1] who derives recurrence and transience criteria in the case where $\ell g(x)$ and $\sigma^2(x)$ asymptotically behave like certain powers of ℓx.

To make this intuition precise let us introduce some notation. Each vector $x \in \mathbb{R}^d$ can be uniquely dissected into two parts

$$x = \hat{x} + \check{x}$$

such that

$$\hat{x} \in \{\lambda r : \lambda \in \mathbb{R}\} \text{ and } \ell\check{x} = 0 .$$

From $\hat{x} = \lambda r$ it follows $\ell x = \lambda \ell r = \lambda$ because of $\ell r = 1$. Thus

$$\hat{x} = r\ell x .$$

Now we may characterize the vectors x close to the ray $\bar{r}$ by the property that $\|\check{x}\|$ is small compared to $\|x\|$.

Similarly the process (X_n) can be splitted:

$$X_n = \hat{X}_n + \check{X}_n .$$

We are now ready to formulate our results. The assumptions are analogous to the 1-dimensional case. The condition of near-criticality now reads

$$\|g(x)\| = o(\|x\|) \quad \text{as } \|x\| \to \infty \tag{A1*}$$

and the condition of moment boundedness gets the following form: There are $c > 0$ and $\delta > 0$ such that for $p = 2 + \delta$ and $\|X_n\| \geq c$

$$\mathbf{E}[\|\xi_n\|^p \mid \mathcal{F}_n] \leq c\sigma^p(X_n) . \tag{A2*}$$

For convenience we formulate here our theorems only for the case that $g(x)$ has only non-negative components, otherwise the required conditions are somewhat more involved.

Theorem 12.3 *Let* (A1*) *and* (A2*) *be fulfilled and let* $g(x) \geq 0$ *(componentwise) for* $\|x\|$ *sufficiently large. Assume that there is an* $\varepsilon > 0$ *and that for any* $a > 0$ *there is some* $b > 0$ *such that for all* $x \in \mathbb{R}^d_+$ *we have*

$$\|x\| \geq b\,, \|\check{x}\|^2 \leq a\|x\| \cdot \|g(x)\| \quad \Rightarrow \quad \ell x \cdot \ell g(x) \leq \frac{1-\varepsilon}{2}\sigma^2(x) . \tag{12.5}$$

Then

$$\mathbf{P}(\|X_n\| \to \infty \text{ } as \text{ } n \to \infty) = 0 .$$

Note that in view of (A1*) the requirement $\|\check{x}\|^2 \leq a\|x\| \cdot \|g(x)\|$ indeed defines a region within $\mathbb{R}_+^d$ which is located in the vicinity of the ray $\overline{r}$ (since $x \in \overline{r}$ implies $\check{x} = 0$). Note also that in the 1-dimensional case the requirement (12.5) reduces to (12.1), because then $\check{x} = 0$ and $\ell x = x$ for any $x \in \mathbb{R}_+$.

Theorem 12.4 *Let (A1*) and (A2*) be fulfilled and let $g(x) \geq 0$ for $\|x\|$ sufficiently large. Also let $\sigma^2(x)$ be bounded away from zero for all $x \in \mathbb{R}_+^d$ with $u < \ell x < v$, where $0 < u < v < \infty$, and let*

$$\sigma^2(x) = O(\|x\|^2 \log^{-2/\delta} \|x\|) \text{ for } x \to \infty \text{ and for some } \delta > 0 \, . \tag{12.6}$$

Assume that there is an $\varepsilon > 0$ and that for any $a > 0$ there is some $b > 0$ such that for all $x \in \mathbb{R}_+^d$ we have

$$\|x\| \geq b \, , \|\check{x}\|^2 \leq a\sigma^2(x) \quad \Rightarrow \quad \ell x \cdot \ell g(x) \geq \frac{1+\varepsilon}{2}\sigma^2(x) \, .$$

Then there is a number $m < \infty$ such that

$$\mathbf{P}(\limsup_n \|X_n\| \leq m \text{ or } \lim_n \|X_n\| = \infty) = 1 \, .$$

If additionally for every constant $c > 0$ there is a natural number n such that we have $\mathbf{P}(\|X_n\| > c) > 0$, then

$$\mathbf{P}(\|X_n\| \to \infty \text{ for } n \to \infty) > 0$$

and

$$\mathbf{P}(\|\check{X}_n\| = o(\|X_n\|) \mid \|X_n\| \to \infty) = 1.$$

Now it is the condition $\|\check{x}\|^2 \leq a\sigma^2(x)$ which in view of (12.6) defines a vicinity of the ray $\overline{r}$. The last statement says that the process diverges in the direction of the ray $\overline{r}$. Again for $d = 1$ this Theorem reduces completely to Theorem 12.2.

Note that we did not specify which norm $\|\cdot\|$ we used. The choice makes no difference because as is well-known all norms on $\mathbb{R}^d$ are equivalent in the sense that for two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ there are number $c_1, c_2 > 0$ such that $c_1\|\cdot\|_1 \leq \|\cdot\|_2 \leq c_2\|\cdot\|_1$. Thus the formulated conditions and statements do not depend on the choice of the norm.

The proof of both theorems will be given elsewhere.

12.5 Example: The Multivariate GW-Process

In the multivariate Galton Watson process $(X_n)_{n\geq 0}$ each generation consists of d different types of individuals, thus $X_n = (X_{n,1}, \dots, X_{n,d})^T$. Then

$$X_{n+1} = \sum_{j=1}^{d} \sum_{i=1}^{X_{n,j}} \zeta_{nij} \text{ or } X_{n+1,k} = \sum_{j=1}^{d} \sum_{i=1}^{X_{n,j}} \zeta_{nijk}$$

with $\zeta_{nij} = (\zeta_{nij1}, \dots, \zeta_{nijd})^T$. Here ζ_{nijk} is considered to be the offspring number of individuals of type k born by the i-th individual of type j in generation n. It is assumed that for $n \geq 0$ and given $X_0, \dots, X_n$ the random vectors ζ_{nij}, $i, j \geq 1$, are independent with a distribution which may depend on j but neither on i nor on n. We allow that this distribution depends also on X_n, then $(X_n)_{n\geq 0}$ is a Markov chain and constitutes a population size dependent multivariate Galton-Watson process as introduced by Klebaner, see e.g. [11]. For a model where M is no longer primitive, compare Jagers and Sagitov [8].

Now given the state $x = (x_1, \dots, x_d)^T \in \mathbb{R}_+^d$ expectations are determined as $\mathbf{E}_x[X_1] = \sum_{j=1}^d \mathbf{E}_x[\zeta_{01j}]x_j$ or

$$\mathbf{E}_x[X_1] = E_x x$$

with the $d \times d$ matrix of expectations

$$E_x = (\mathbf{E}_x[\zeta_{011}], \dots, \mathbf{E}_x[\zeta_{01d}]) \, .$$

In our context this means that

$$g(x) = (E_x - M)x$$

with some primitive matrix M and

$$\xi_n = \sum_{j=1}^{d} \sum_{i=1}^{X_{n,j}} (\zeta_{nij} - \mathbf{E}_x[\zeta_{01j}]) \, .$$

By conditional independence we get

$$\sigma^2(x) = \mathbf{E}_x[(\ell \xi_0)^2] = \sum_{j=1}^{d} \mathbf{Var}_x(\ell \zeta_{01j}) x_j$$

or

$$\sigma^2(x) = \ell \Gamma_x \ell^T$$

with

$$\Gamma_x = \sum_{j=1}^{d} \mathbf{Cov}_x(\zeta_{01j}) x_j$$

and the $d \times d$ covariance matrices $\mathbf{Cov}_x(\zeta_{01j})$ of $\zeta_{01j}, j = 1, \ldots, d$.

As to condition (A2*) it is suitable here to work with the ℓ_1-norm $\|x\| = \sum_k |x_k|$. Then for $p = 2 + \delta > 2$

$$\|\xi_n\|^p = \Big(\sum_{k=1}^{d} \Big|\sum_{j=1}^{d} \sum_{i=1}^{X_{n,j}} (\zeta_{nijk} - \mathbf{E}_x[\zeta_{01jk}]\Big|\Big)^p \le d^{2p} \sum_{j,k=1}^{d} \Big|\sum_{i=1}^{X_{n,j}} (\zeta_{nijk} - \mathbf{E}_x[\zeta_{01jk})]\Big|^p$$

Applying again the Marcinkiewicz-Zygmund and the Hölder inequality we obtain, for some $c > 0$,

$$\mathbf{E}_x[\|\xi_0\|^p] \le cd^{2p} \sum_{j,k=1}^{d} x_j^{p/2} \mathbf{E}_x[|\zeta_{01jk} - \mathbf{E}_x[\zeta_{01jk}]|^p] \, .$$

Assuming now that there is a number $b > 0$ such that for all $x \in \mathbb{R}_+^d$

$$\mathbf{E}_x[|\zeta_{01jk} - \mathbf{E}_x[\zeta_{01jk}]|^p] \le b\mathbf{E}_x[(\zeta_{01jk} - \mathbf{E}_x[\zeta_{01jk}])^2]^{p/2} \tag{12.7}$$

we obtain

$$\mathbf{E}_x[\|\xi_0\|^p] \le bcd^{2p} \Big(\sum_{j,k=1}^{d} x_j \mathbf{E}_x[(\zeta_{01jk} - \mathbf{E}_x[\zeta_{01jk}])^2] \Big)^{p/2} \, ,$$

that is

$$\mathbf{E}_x[\|\xi_0\|^p] \le bcd^{2p} (\text{trace } \Gamma_x)^{p/2} \, .$$

Thus to attain validity of (A2*) we require besides (12.7) that there is a constant $c > 0$ such that for all $x \in \mathbb{R}_+^d$ we have

$$\text{trace } \Gamma_x \le c \, \ell \Gamma_x \ell^T \, .$$

This is fulfilled if e.g. the covariance matrices $\mathbf{Cov}_x(\zeta_{01j})$ have only non-negative entries, but it may fail in general.

Now we are ready to apply our theorems to special cases as those discussed by Klebaner [11] and Adam [1]. Details are left to the reader.

References

1. Adam, E.: Criterion for unlimited growth of critical multidimensional stochastic models. Preprint, arXiv 1502.04046 [math.PR] (2015)
2. Albeverio, S., Kozlov, M.V.: On recurrence and transience of state-dependent branching processes in random environment. Theory Probab. Appl. **48**, 575–591 (2004)
3. Daley, D., Hull, D.M., Taylor, J.M.: Bisexual Galton-Watson branching processes with superadditive mating functions. J. Appl. Probab. **23**, 585–600 (1986)
4. González, M., Martínez, R., Mota, M.: On the unlimited growth of a class of homogeneous multitype Markov chains. Bernoulli **11**, 559–570 (2005)
5. González, M., Molina, M., del Puerto, I.: Asymptotic behaviour of critical controlled branching processes with random control functions. J. Appl. Probab. **42**, 463–477 (2005)
6. González, M., Martínez, R., Mota, M.: Rates of growth in a class of homogeneous multidimensional Markov chains. J. Appl. Probab. **43**, 159–174 (2006)
7. Granger, C., Inoue, T., Morin, N.: Nonlinear stochastic trends. J. Econ. **81**, 65–92 (1997)
8. Jagers, P., Sagitov, S.: The growth of general population-size-dependent branching processes year by year. J. Appl. Probab. **37**, 1–14 (2000)
9. Kersting, G.: On recurrence and transience of growth models. J. Appl. Probab. **23**, 614–625 (1986)
10. Klebaner, F.: On population-size-dependent branching processes. Adv. Appl. Probab. **16**, 30–55 (1984)
11. Klebaner, F.: Asymptotic behavior of near-critical multitype branching processes. J. Appl. Probab. **28**, 512–519 (1991)
12. Küster, P.: Generalized Markov branching processes with state-dependent offspring distributions. Z. Wahrscheinlichkeit **64**, 475–503 (1983)
13. Molina, M., Mota, M., Ramos, A.: Bisexual Galton-Watson branching process with population–size–dependent mating. J. Appl. Probab. **39**, 479–490 (2002)
14. Molina, M., Jacob, C., Ramos, A.: Bisexual branching processes with offspring and mating depending on the number of couples in the population. Test **17**, 265–281 (2008)
15. Molina, M., Mota, M., Ramos, A.: Some contributions to the theory of near-critical bisexual branching processes. J. Appl. Probab. **44**, 492–505 (2007)
16. Seneta, E.: Non-negative Matrices and Markov Chains, 2nd edn. Springer, New York (1981)

Chapter 13
The Weighted Branching Process

Uwe Roesler

Mathematics Subject Classification (2000): Primary 60F05, Secondary 37C25, 60F17

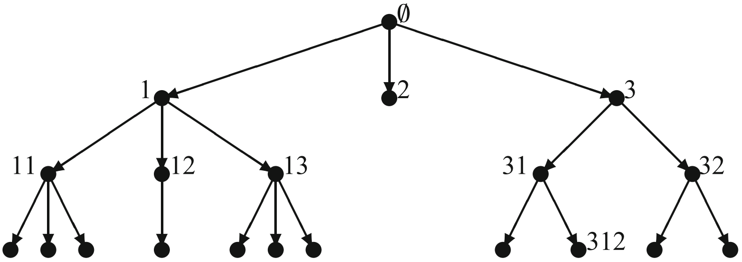
13.1 Weighted Branching Process

Branching processes have a lot of applications in the real world. Basically, one counts the number of offspring or individuals in a certain generation. But that is only part of the story. In reality the mother (parents) passes some value to the children and they to their children and so on. Examples of values are the DNA sequence, money, knowledge, influence and so on.

We will present a general suitable mathematical model, the Weighted Branching Process (WBP) [37, 39] $(V, L, G, *)$.

U. Roesler (✉)
Mathematisches Seminar, Christian-Albrechts-Universität zu Kiel, Ludewig-Meyn-Strasse 4, 24098 Kiel, Germany
e-mail: roesler@math.uni-kiel.de

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_13



Picture of a tree

For the tree we use the Ulam-Harris notation. Let $V = \mathbb{N}^* = \cup_{n=0}^{\infty} \mathbb{N}^n$ be the set of finite sequence of integers including the empty sequence called $\emptyset$. We use $|v| = n$ for $v = (v_1, v_2, \dots, v_n) = v_1 v_2 \dots v_n$ suppressing the root $\emptyset$. V is a (genealogical) graph with directed edges $(v, vi),\ v \in V,\ i \in \mathbb{N}$.

$(G, *)$ is a measurable monoid with a grave. (A monoid is a semi group with a neutral element ($e = e * g = g * e$ for all $g \in G$). A grave is an element $\triangle \in G$ such that $\triangle = \triangle * g = g * \triangle$.)

Each edge (v, vi) carries as weight $L_{v,vi}$ a random variable with values in G. Define on every path (v, vw) a weight (length) $L_{v,vw}$ recursively by

$$L_{v,vwi} = L_{v,vw} * L_{vw,vwi}$$

For notational reasons we use also (v, v) as path and define $L_{v,v}$ as the neutral element e.

The object $(V, L, G, *)$ is called Weighted Branching Process (WBP) if $(L_{v,vi})_{i \in \mathbb{N}},\ v \in V$ are iid rvs. For specified edge weights we obtain a random dynamical system indexed by a tree

$$L_{v,x} = L_{v,w} * L_{w,x}$$

$v \preceq w \preceq x$ in genealogical order. Sometimes we specify only $(L_i)_{i \in \mathbb{N}}$ or the distribution of it. We then have to choose iid rvs in order to obtain a WBP. More precisely we face a Markov chain indexed by a tree with transition kernel $G \times \sigma(G^{\mathbb{N}}) \ni (x, A) \mapsto P((x * L_i)_i \in A)$. On the other hand, given a transition kernel on $G \times \sigma(G^{\mathbb{N}})$, under very weak assumptions a dynamical system with these transitions [10] exists. Our objects of interest will always be WBP even if we specify only the distribution.

We suppress the root $\emptyset$ and grave valued objects whenever possible. The picture above shows all v such that $L_{\emptyset,v} =: L_v \neq \triangle$. These L_v are observable. We keep the common interpretation of offspring, descendant, ancestor, child, mother and so on.

Ex: **Free Semi group**: If L_i are some fixed element of G then $L_i,\ i \in \mathbb{N}$, generate $G = \{L_v \mid v \in V\} \cup \{e, \triangle\}$ as free semi group adding the neutral element and the grave.

Here are some examples in the reals.

Ex: **Bienaymé-Galton-Watson (BGW):** [9] Take $G = \{0, 1\}$ with multiplication. 0 is the grave. Then the BGW process $Z_n = \sum_{|v|=n} L_v$ counts the number of offspring in n-th generation. $m = E \sum_{i\in\mathbb{N}} L_i$ is the average offspring.

Ex: **Branching Random Walk (BRW):** $G = [0, \infty)$ with multiplication. Interesting objects are for example $Z_n = \sum_{|v|=n} L_v$ and $\sup_{|v|=n} L_v$. In his Ph.D. and a series of papers Biggins [14] introduced the branching random walk on $(\mathbb{R}, +)$ and considered the asymptotic of the sum of weighted positions in n-th generation and the rightmost particle. Taking the exponential function his setting corresponds to a WBP on the positive reals with multiplication.

These examples can be extended to G the reals or complex numbers with multiplication or to Hilbert spaces and so on. They include multi type branching processes [9] or multi type branching random walk [15].

The interpretation of the above examples is as follows: One individual branches and moves the children. Mathematically we interpret $g \in \mathbb{R}$ as a map $T_g : \mathbb{R} \to \mathbb{R}, T_g(x) = xg$ and consider $L_v(x)$ as the value of the particle v starting the process at the root with a particle of weight x. In general, without loss of generality, G is a subset of some function space H^H (always $G = H$ does the job) with composition, $x * y = x \circ y$ or $x * y = y \circ x$. The algebraic description is by G acting right on $H = H_l$ via $*_l : H_l \times G \to H_l$ or left on $H = H_r$ via $*_r : G \times H_r \to H_r$. (Notice the formal definition of a semi group acts right or left and our choice of the symbols $H_l, H_r, *_l, *_r$ inspired by the natural position of H.) Then $T_g(x) = x *_l g$ or $g *_r x$. The symbol $*_l$ corresponds to the interpretation of a particle starting at the root. The symbol $*_r$ is connected to some values at the boundary of V at ∞, see examples coming up. Also a combination $x *_l g *_r y$ with values at the boundary of V is natural (with the right formal modification).

The next examples concern self similar sets and random affine maps on vector spaces. (Affine maps are not commutative.) It is the celebrated world of fractals [11, 12].

Ex: **Cantor set:** Consider the iteration

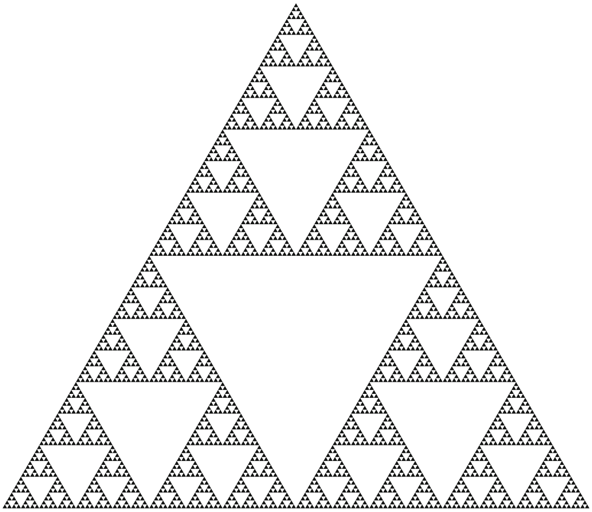
$$X_n = \frac{1}{3}X_{n-1} + \frac{2}{3}B_n,$$

where $B_n, \ n \in \mathbb{N}$, are independent Bernoulli distributed rvs with parameter $1/2$. Plotting the sequence $(X_n)_n$ for the starting point $x_0 = 0$ (or any) provides (asymptotically) a picture of the Cantor set (explanation later). The sequence arises from a somewhat degenerate WBP, G is generated by the affine maps $\mathbb{R} \ni x \mapsto g_1(x) = \frac{x}{3}$ and $\mathbb{R} \ni x \mapsto g_2(x) = \frac{x+2}{3}$ with composition $x * y = y \circ x$ and enlarged by the identity and the grave. G acts right on the reals $H = \mathbb{R}$. Let $Y_v, v \in V$ be independent Ber$(1/2)$ rvs and put $L_{v,v1} = Y_v g_1 + (1 - Y_v) g_2$ and $L_{v,vi}$ the grave for $i \geq 2$. Only vertexes v consisting purely of 1's get a non grave weight.

Cantor set

For another description let G and H be as above. Define $L_{v,vi} = g_i$ for $i = 1,2$ and the grave otherwise. Consider the range $K_n := \{Z_n(x) \mid x \in [0,1]\}$ of $Z_n = \sum_{|v|=n} L_v$. Then $K := \cap_n K_n$ is the Cantor set.

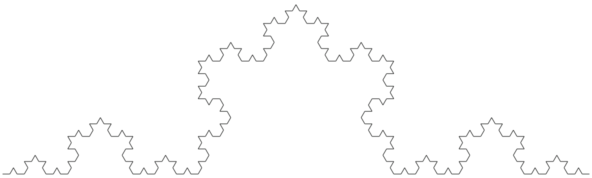
There are more interesting (deterministic) self similar examples. One is the Sierpinsky triangle



Sierpinski triangle

$X_n = A_n X_{n-1} + B_n$ in $\mathbb{R}^2$. Here A_n is the matrix $A := \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$ and B_n are iid with values $(0,0)$, $(1/4, 1/2)$, $(1/2, 0)$ taken with probability $1/3$.

We mention also the Koch curve without explanation.



Koch curve

Ex: **Fern:** Here is a picture of my favorite, the fern.

Fern

Draw the $X_n, n \in N$ with values in $\mathbb{R}^2$ satisfying the recursion $X_n = A_n X_{n-1} + B_n$ with (A_n, B_n) iid. The distribution of the random matrix A and the random vector is

$$\left(\begin{pmatrix} 0.02 & 0 \\ 0 & 0.27 \end{pmatrix}, \begin{pmatrix} 150 \\ 0 \end{pmatrix}\right),$$

$$\left(\begin{pmatrix} -0.139 & 0.263 \\ 0.246 & 0.224 \end{pmatrix}, \begin{pmatrix} 171 \\ -10.8 \end{pmatrix}\right),$$

$$\left(\begin{pmatrix} 0.16 & -0.215 \\ 0.222 & 0.176 \end{pmatrix}, \begin{pmatrix} 122.4 \\ 26.79 \end{pmatrix}\right),$$

$$\left(\begin{pmatrix} 0.781 & 0.034 \\ -0.032 & 0.739 \end{pmatrix}, \begin{pmatrix} 32.25 \\ 81 \end{pmatrix}\right),$$

with probability $0,01; 0,15; 0,13; 0,71$. An explanation for such phenomena (even in Hilbert spaces) is given in [16, 24].

Another big source of WBP examples arises by interval splitting.

Ex: **Interval splitting:** Consider the set G of intervals $[a, b),\ 0 \le a \le b \le 1$, Define the operation $*$ on intervals by

$$[a, b) * [c, d) := [a + c(b - a), a + d(b - a)).$$

We will use $I_{v,vi}$ and I_v in the sequel instead of $L_{v,vi}, L_v$ as before. Notice $g \in G$ is identified with the map $T_g(x) = x * g$ and $H_l = G$. We use for simplicity right half open intervals, but open, closed or left half open intervals are possible.

For the Cantor set let $g_1 := [0, 1/3], g_2 := [2/3, 1]$ and let G be generated by g_1, g_2 under $*$ adding the neutral element and the grave. Put $I_{v,vi} = g_i$ for $i = 1, 2$

and the grave otherwise. Then

$$Z_n := \cup_{|v|=n}[0,1] * I_v = \bigcup_{m=0}^{\frac{3^n-1}{2}} \left[\frac{2m}{3^n}, \frac{2m+1}{3^n}\right]$$

is the K_n from above and provides in the limit the Cantor set $K = \cap_n Z_n = \{x = \sum_{i=1}^{\infty} \frac{x_i}{3^i} \mid x_i \in \{0,2\}\}$.

Ex: ***b*-ary splitting:** Split $[0,1)$ into b intervals of equal length. Then $I_i = [\frac{i-1}{b}, \frac{i}{b})$ for $i \leq b$. Notice $I_v,\ |v| = n$, is a partition of $[0,1)$ for any $n \in \mathbb{N}$.

The next step of generality is a random partitioning $I_i,\ i \in \mathbb{N}$ of the unit interval $[0,1)$. A self similar continuation (=a WBP on intervals) provides $I_v, v \in V$. Define another WBP $(V, L, G, *)$ on the reals $(G = \mathbb{R})$ with multiplication. Objects of interest are processes like $[0,1) \ni t \mapsto Z_n(t) := \sum_{|v|=n} 1\!\!1_{I_v}(t) L_v$ or functions of it.

For the Cantor set take a ternary splitting $I_1 = [0, 1/3],\ I_2 = (1/3, 2/3),\ I_3 = [2/3, 1]$ and a positive WBP with $L_1 = |I_1|,\ L_3 = |I_3|$ and 0 otherwise. Then the α satisfying $E\sum_i L_i^\alpha = 1$ is the Hausdorff dimension ($\alpha = \frac{\ln 2}{\ln 3}$) and the function $\mathbb{R} \supset A \mapsto \lim_{n\to\infty} \int_0^A (Z_n(t))^{\alpha-1} dt$ the Cantor measure (=α-Hausdorff measure restricted to the Cantor set [18]). The Cantor set is the support of the Cantor measure.

Ex: **Mandelbrot Cascades:** Let $I_v,\ v \in V$, be an b-ary interval splitting as above. Further take a WBP on $G = [0, \infty)$ with multiplication and weights, such that L_i is the grave for $i > b$. Let $\mathcal{F}_n$ be the σ-field generated by $L_{v,vi}, |v| < n, i \in \mathbb{N}$ and $\mathcal{G}_n$ be the σ-field generated by $I_v, |v| = n$. The sequence Q_n of measures $Q_n(A) = \sum_{|v|=n} L_v \sigma(A)$ on $\mathcal{G}_n$ are called Mandelbrot cascades for the measure σ. Under suitable assumptions [23, 26] the random measures Q_n converge to some random measure Q_∞ on the Borel σ-field of the boundary of V at infinity.

Under the Lebesgue measure σ and the assumption $E\sum_i L_i = b$ is $(Q_n)_n$ a martingale measure, $E(\int f dQ_{n+1} \mid \mathcal{F}_n) = \int f dQ_n$ for all $f \geq 0$ measurable with respect to $\mathcal{G}_n$ [23]. Further $Z_n(t) = b^{-n} \sum_{|v|=n} L_v 1\!\!1_{I_v}(t)$ is a martingale in n for every $t \in [0,1)$ [23]. There are several extensions for different σ and weakening the assumptions leading to stochastic fixed point equations (SFE) on the reals (see section on SFE) [17, 21, 23, 25, 26, 41] or to SFE for processes (see last section) [13]. Most of them (if not all) are WBP and the limits appearing there are stochastic fixed points.

The next example is from computer science, sorting and searching [38].

Ex: **Binary Search Tree:** Let H be the set of infinite sequences of different reals. Define $g_1 : H \to H$ by $g_1(x) = (x_{i_1}, x_{i_2}, \ldots)$ where $i_1 := \inf\{n > 1 \mid x_n < x_1\}$ and $i_j := \inf\{n > i_{j-1} \mid x_n < x_1\}$. $g_1(x)$ is the subsequence of reals strictly smaller than x_1. Define g_2 analogue replacing the $<$ by $>$. Let G be generated by g_1, g_2 including the identity and the grave. Put $L_i = g_i$ for $i \leq 2$ and otherwise the grave. Let $\Phi(v)$ be the first element in the vector $L_v(x)$. Then $T_n(x) := \{v \in V \mid \Phi(v) \in \{x_1, x_2, \ldots, x_n\}\}$ together with the map $T_n(x) \ni v \to \Phi(v)$ is called the binary tree (for $x_1, x_2, \ldots, x_n$). Traversing from left to right through the tree $T_n(x)$ provides the reals $x_1, x_2, \ldots x_n$ in natural order.

Ex: **Phylogenetic trees:** Given is the DNA-sequence of some individuals or species living today. Find the phylogenetic tree of them [32]. The idea is, that mutations appears seldom, but regularly. Model the mutations by a Poisson process and take the Maximum-Likelihood estimator over all trees including mutations. These empirical phylogenetic trees are reasonable good estimates close to estimates with other methods.

13.2 Stochastic Fixed Point Equation

In general stochastic fixed point equations (SFE) are of the form

$$X \stackrel{\mathrm{d}}{=} \varphi((X_i)_i).$$

Here $\varphi, X_i,\ i \in \mathbb{N}$, are independent and all X_i have the same distribution as X.

Ex: **Max type:** Take φ, a random map with values in $\mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, as the supremum $\varphi((x_i)) = \sup_i x_i + B$, where B is a rv [7, 40].

We consider here only affine maps φ on some vector space, i.e. SFE of sum type

$$X \stackrel{\mathrm{d}}{=} \sum_{i \in \mathbb{N}} A_i X_i + B, \tag{13.1}$$

whenever the above is well defined. Here the rvs $((A_i)_i, B),\ X_j,\ j \in \mathbb{N}$ are independent and the X_j have all the same distribution as X. The distribution of $((A_i)_i, B)$ is given.

Many important distributions satisfy such a SFE in the reals.

Ex: **Gauss:** The equation $X \stackrel{\mathrm{d}}{=} \frac{X_1}{\sqrt{2}} + \frac{X_2}{\sqrt{2}}$ is uniquely solved by a Gauss distribution $N(0, \sigma^2)$ for some $0 \leq \sigma^2 < \infty$.

Ex: **Cauchy:** The equation $X \stackrel{\mathrm{d}}{=} \frac{X_1}{2} + \frac{X_2}{2}$ is solved by a symmetric Cauchy(b) distribution, (density $x \mapsto \frac{b}{\pi(b^2+x^2)}$) and by constants.

Ex: **α-stable-distribution:** α-stable distributions in $\mathbb{R}$ are the limits of linear normalized sums of iid rvs. Alternatively X is α-stable distributed iff $X + c \stackrel{\mathrm{d}}{=} aX_1 + bX_2$ for all $a, b \in \mathbb{R}$ satisfying $|a|^\alpha + |b|^\alpha = 1$ and some $c \in \mathbb{R}$. What happens if we have the above SFE only for some specific a, b, c? Are the solutions stable [6]?

Ex: **BGW:** $\frac{Z_n}{m^n}$ converges a.e. to W and W satisfies

$$W \stackrel{\mathrm{d}}{=} \sum_{i \leq N} \frac{1}{m} W_i,$$

N the number of offspring.

Ex: **Quicksort:** The running time X of the divide and conquer algorithm Quicksort, correctly normalized [35], satisfies

$$X \overset{\mathrm{d}}{=} UX_1 + (1-U)X_2 + C(U). \tag{13.2}$$

Here U is uniformly distributed and $C(x) = 2x\ln x + 2(1-x)\ln(1-x) + 1$. The solution is unique in the set L^1, but there are more solutions [19] of (13.2).

The last example motivated the study of SFE as objects of its own interest. Neininger [28] in a series of papers analyzed many more stochastic divide-and-conquer algorithms, ending up with SFEs.

- Why are we interested solving (13.1) in distribution? There always exists rvs on a common probability space satisfying

  $$X = \sum_i A_i X_i + B.$$

  The quick answer is, distributions are on lower level than rvs. Every rv has a distribution, but for given distribution there are many rvs with that distribution. More to come.
- Why are solutions of (13.1) called fixed point?
  Consider the map $K = K_\nu$ from distribution to distributions

  $$K(\mu) \overset{\mathrm{d}}{=} \sum_i A_i Y_i + B,$$

  $((A_i)_i, B)$ has distribution ν and is independent of all $Y_i \overset{\mathrm{d}}{=} \mu$. Fixed points of K are exactly the solutions of the SFE. Methods for dealing with K are e.g. the contraction method [39] using the Wasserstein metric [36, 39] or Zolotarev metric introduced by Neininger [28], Neininger and Rüschendorf [30]. If $B = 0$ then K is called the smoothing transform [17, 21, 23].
- Where is the dynamic and connection to weighted branching processes?
  Iterate the homogeneous SFE

  $$X = \sum_i A_i X_i + B = \sum_i \sum_j A_i A_{ij} X_{ij} + \sum_i A_i B_i + B = \dots$$

  This leads to a WBP, a genealogical tree (with possible infinite branches), the mother passes her value randomly transformed to her children. We have dependence within a family, but independence for families.

13.3 Forward and Backward Dynamics

Without loss of generality we may take G as a subset of H^H. The $*$ is the composition, but be aware of the order!

Forward Dynamics (Genealogy) Start with some value $h \in H$ at the root $\emptyset$ and every vertex v gives its weight to her i-th child vi randomly transformed by the map $L_{v,vi}$.

Formal: Let H_l be a set and G acts right on H_l via $*_l : H_l \times G \mapsto H_l$. We obtain $h *_l L_v$ as the weight of v for starting weight h at the root. h might also be a rv independent of everything else.

Ex: For G the reals with multiplication we usually identify a real number a with the transformation $\mathbb{R} \ni x \mapsto T_a(x) = ax$. Therefore for the BGW starting with one particle of weight 1 at the root we obtain the value $1 *_l L_v = L_v$ which is either 0 or 1 with the interpretation dead or alive.

Ex: The descriptions of the Cantor set is the forward dynamic, starting with points in the unit interval or the unit interval itself.

Backward Dynamics Now we look backwards in genealogical time. Give the vertex v a weight C_v, take it back to the root $L_v *_r C_v$ and consider the total weight $R = \sum_{v \in V} L_v *_r C_v$. Another procedure is, give vertexes v in n-th generation some values C_v and define recursively $C_w := \sum_{i \in \mathbb{N}} L_{w,wi} *_r C_{wi},\ |w| < n$. Go all the way back to the root and then take the limit $n \to \infty$. Intuitively, the limit is like putting some weight on the boundary of V and going back. More general, let the children determine the weight of their mother and put some starting weight on the boundary at infinity.

Formal: Let H_r be a set, G acts left on H_r via $*_r : G \times H_r \mapsto H_r$ and there is a (random) function $\psi : H_r^{\mathbb{N}} \to H_r$. Put some values $C_v \in H_r$ on the n-th generation $|v| = n$. Then on the $n-1$-th generation $|w| = n-1$ we have the weight $w \mapsto \psi((L_{w,wi} *_r C_v)_i) \in H_r$. Continue this procedure up to the root. The standard assumption on the input is $((L_{v,vi})_i, C_v),\ v \in V$, are iid rvs.

Ex: As an example take the SFE

$$X \overset{\mathrm{d}}{=} \frac{X_1}{\sqrt{2}} + \frac{X_2}{\sqrt{2}} \tag{13.3}$$

uniquely solved by a Gauss distribution $N(0, \sigma^2)$ for some $0 \le \sigma^2 < \infty$. Take a WBP on $\mathbb{R}$ with $L_i = 1/\sqrt{2}$ for $1 \le i \le 2$ and 0 otherwise. Put some independent $N(0, \sigma^2)$ rvs $C_v,\ |v| = n$ on the n-th generation, and obtain $Y^v = \sum_{|vw|=n} L_{v,vw} *_r C_{vw}$ for all $|v| < n$. All $Y^v,\ |v| < n$ for the tree vV are $N(0, \sigma^2)$ distributed. But if we try to extend n to infinity and obtain as limit a normal distribution, there is the difficulty, that in the limit one needs values at the boundary of V at infinity.

Ex: Consider a BGW process and W the limit of the martingale $\frac{Z_n}{m^n}$, $m = EZ_1$. Then W satisfies

$$W = \sum_{i\in\mathbb{N}} \frac{L_i}{m} W^i = \sum_{|v|=n} \frac{L_v}{m^{|v|}} W^v, \tag{13.4}$$

where W^v is the martingale limit on the tree vV.

Endogenous or Non endogenous We call solutions of a SFE endogenous [1], if they are measurable with respect to the corresponding WBP. Otherwise they are called non endogenous. The solution W in example (13.4) is measurable as limit of a martingale and solutions to (13.3) are not. For endogenous solutions we can easily give a version with equality rather than in distribution simultaneous for all subtrees vV. For non endogenous there exists such versions, but we have to put some values at the boundary of V at infinity. This requires some techniques.

The forward and backward dynamics is connected to the forward and backward view. Just for the case of simplicity we take G the reals with multiplication and grave 0. We take the genealogical dynamics. Our object of interest is

$$Z_n := \sum_{|v|=n} L_v.$$

The **Forward** and **Backward view** is in analogy to Markov processes.

The **forward** equation is

$$Z_n = \sum_{|v|=n-1} L_v \sum_i L_{v,vi}.$$

The forward view is on random variables. The forward structure is (often) a martingale

$$W_n := \frac{Z_n}{m^n} \longrightarrow_n W,$$

$m := E\sum_i L_i$. The forward result is a.e. convergence (probability theory).

The **backward** equation is

$$Z_n = \sum_{i\in\mathbb{N}} L_i Z^i_{n-1},$$

where $Z^i_{\cdot}$ is for the tree iV. The backward view is on measures. The backward structure is an iteration on $K = K_\nu$, $K(\mu) \overset{\mathcal{D}}{=} \sum_i L_i X_i$. The backward result is weak convergence (measure theory)

$$K^n(\mu_0) \to_n \mu = K(\mu). \tag{13.5}$$

The connection between both is $K^n(\delta_1) \overset{\mathcal{D}}{=} Z_n$.

Analogous considerations hold for

$$R_n = \sum_{|v|<n} L_v *_r C_v,$$

$((L_{v,vi})_i, C_v),\ v \in V$, iid, $C_v : \Omega \to G$. The backward view is

$$R_n = \sum_i L_i *_r R^i_{n-1} + C.$$

With forward tools like $R_n \to R$ point wise we obtain

$$R = \sum_i L_i *_r R^i + C.$$

In case $R_n \to R$ only converges in distribution we obtain only equality in distribution.

13.4 SFE of Sum Type

In this section we consider solutions of (13.1) in the reals. Especially we are interested in **all** solutions. We distinguish between the homogeneous sum type, B is identically 0, and the non homogeneous sum type.

Let us start with the homogeneous sum type

$$X \overset{\mathrm{d}}{=} \sum_i A_i X_i \tag{13.6}$$

Consider a WBP on the reals $G = \mathbb{R}$ with multiplication and the $(L_i)_i$ are distributed like $(A_i)_i$. Then $\frac{Z_n}{m^n}$, $Z_n := \sum_{|v|=n} L_v$, $m = E\sum_i L_i$ is under suitable assumptions a martingale relative to the filtration $\mathcal{A}_n$ generated by $L_{v,vi}, |v| < n, i \in \mathbb{N}$. If we consider positive weights L_i and $0 < m < \infty$ then the positive martingale $\frac{Z_n}{m^n}$ converges a.e. to some W and W is an endogenous solution of the SFE $W \overset{\mathrm{d}}{=} \sum_i \frac{L_i}{m} W_i$. W might be identical 0, which is always a solution. Here is a Kesten-Stigum analogue [9, 14, 27] deciding on $W \equiv 0$. Define the function $(0,\infty) \ni \alpha \mapsto m(\alpha) = E\sum_i |L_i|^\alpha \in [0,\infty]$. We use $m = m(1)$.

Theorem 13.1 (Biggins) *Let $(G, *)$ be $[0,\infty)$ with multiplication, $m < \infty$ and $m'(1)$ exists and is finite. Then are equivalent*

i) $P(W = 0) < 1$
ii) $EW = 1$
iii) $E(Z_1 \ln^+ Z_1) < \infty$ *and* $m'(1) < m \ln m$

$P(W=0)$ is either 1 or the extinction probability of the corresponding BGW. Some results can be extended to real valued factors L_i.

Usually non endogenous solutions are harder to obtain, [4–6, 8, 17, 21–23, 25]. A final (almost) complete answer is a consequence of the Ph.D.-thesis of Meiners [8]. We treat only the the positive case, $G=[0,\infty)$. Let S be the set of all decreasing left continuous functions $f:\mathbb{R}_+\to[0,1]$ satisfying $f(0)=1$ and the fixed point equation

$$f(t)=E\prod_i f(tL_i) \tag{13.7}$$

for all t. This problem corresponds to the SFE

$$Y \overset{\mathcal{D}}{=} \inf_i \frac{Y_i}{L_i} \tag{13.8}$$

of infimum type. By convention put $\inf\emptyset=\infty=\frac{x}{0}=\infty$ for every $x\in G$. The correspondence is given by the survival function $f(t)=P(Y\geq t)$. The connection to the SFE (13.6) is: The Laplace transform $\varphi(t)=E(e^{-tX})$ of a solution X of the SFE (13.6) is an element of S.

Here are some assumptions. A1) The smallest multiplicative closed group generated by the support of all L_i, $i\in\mathbb{N}$ is G.

A2) $P(\sum_i \mathbb{1}_{L_i>0}<\infty)=1$
A3) There is an $\alpha\leq 1$ such that $m(\alpha)=1$ and $m(\alpha)<m(\beta)$ for all $\beta<\alpha$.
A4) There is an $\beta<\alpha$ such that $m(\beta)<\infty$.

Theorem 13.2 (Alsmeyer-Biggins-Meiners) *Assume A1–A4 and* $G=[0,\infty)$. *Then there exists a positive random variable Z not identically* 0 *solving the SFE*

$$Z \overset{\mathcal{D}}{=} \sum_i L_i^{\alpha} Z_i. \tag{13.9}$$

The non degenerate functions in S form exactly the parametrized family

$$f(t)=E(e^{-Zht^{\alpha}}). \tag{13.10}$$

$h\in(0,\infty)$.

If we are interested in solutions X of (13.6) then use: A solution $f\in S$ (13.10) is a Laplace transform iff $\alpha\leq 1$. Then X is of the form $Z^{1/\alpha}Y$ with Y a symmetric α-stable distribution. It is a mixture of symmetric α-stable distributions.

For $G=\mathbb{R}$ a similar result is known [5], solutions are again mixtures of α-stable distributions.

We come now to non homogeneous solutions $X \overset{\mathrm{d}}{=} \sum_i A_iX_i+B$ and ask for **all** solutions. Some endogenous solutions are easy to find via the contraction method [39], a backward approach. See [2] for a recent overview.

Consider the operator $K = K_\nu$ (13.5) on probability distributions. Show K is a strict contraction with respect to some suitable metric and use Banach Fixed Point theorem. Suitable metrics are the Wasserstein [39] or Zolotarev metric [30], both [2] with nice properties fitting the recursive structure,

$$d_p(\mu, \nu) = \inf\{\|X - Y\|_p \mid X \overset{\mathrm{d}}{=} \mu,\ Y \overset{\mathrm{d}}{=} \nu\},$$

$$\xi_p(\mu, \nu) = \sup_f |Ef(X) - Ef(Y)|,$$

where $D^{\lfloor p \rfloor} f$ is Hölder $(p - \lfloor p \rfloor)$-continuous. There are more nice metrics [28].

The forward approach uses martingale theory. Consider a WBP on the reals $G = \mathbb{R}$ with multiplication and take independent rvs $((L_{v,vi})_i, C_v),\ v \in V$ with the same distribution as $((A_i)_i, B)$. $R_n := \sum_{|v|<n} L_v C_v$ is a martingale in case $E(C) = 0$. Under suitable conditions, like L^p-martingale $p > 1$ [3, 30, 39], R_n converges to some R, which is a solution to the SFE (13.1).

Non endogenous Rüschendorf [40] characterized the general solution of the non homogeneous SFE as the sum of one solution of the non homogeneous and the general solution of the homogeneous [4, 22]. Under suitable assumptions the set of **all** solution is given by

$$R + Z^{1/\alpha} Y,$$

where $(R, Z), Y$ are independent, $R = \lim_n R_n, Z$ a solution of $Z \overset{\mathcal{D}}{=} \sum_i L_i^\alpha Z_i$ and Y an α-stable distribution to certain parameters.

In the example Quicksort we obtain R the Quicksort distribution, Z a constant, since $\sum_i L_i$ is identical 1 and Y a symmetric Cauchy distribution, which is 1-stable.

13.5 The Contraction Method

The contraction method, roughly described in the previous section, uses the operator $K = K_\nu$ (13.5). Under a suitable metric [33] $K^n(\mu_0)$ will converge in distribution for some starting measure μ_0 to some fixed point $\mu = K(\mu)$. Examples are the Cantor set description via affine maps, where X_n converges in distribution to the unique invariant probability measure μ and the sequence (X_n) comes close to every point in the support of μ. The SFE is $X \overset{\mathrm{d}}{=} AX + B$ and X_n is ergodic [24] on the support of X respectively μ. In this manner we obtain also a picture for the Sierpinsky triangle and the fern.

The real strength of the contraction method shows up for ‘dirty’ recursions which appear a lot for random divide-and-conquer algorithms

$$X(n) \overset{\mathcal{D}}{=} \sum_i A_i(I(n))X_i(I(n)) + B(I(n)),$$

where $(A_i(\cdot))_i, B(\cdot), I(n)), X_j, j \in \mathbb{N}$ are independent, $I(n) < n$.

Assuming $I(n) \to_n \infty,\ A_i(n) \to_n A_i,\ B(n) \to_n B$ and $X(n) \to X$ in distribution we hope for $X \overset{\mathrm{d}}{=} \sum_i A_i X_i + B$. The first step, usually the easy one, is to solve the SFE, and the second is to show $X(n) \to_n X$ [29, 39] in some nice sense.

This approach led to the first running time analysis of a stochastic divide-and-conquer algorithm [35]. Quicksort still serves as an icon since that time. The input is a list of n different numbers and the output is the ordered list. We present the version with random input (contrary to random pivot). Consider a binary tree and put a sequence $U_1, U_2, \ldots, U_n$ of iid rvs with a uniform distribution to the root. Take the first number U_1 in the list, and put successively any number strictly smaller to the vertex 1 and the others to the vertex 2, keeping the order of appearance. The U_1 remains at the root. Then recall the algorithm as long as any vertex has a list of length at least 2. The final tree of occupied vertexes is the binary search tree $T_n = T_n(U_1, \ldots, U_n)$, see the example.

Let $Y(n)$ denote the total number of comparisons required to sort the input, which has the interpretation of being proportional to the running time of the algorithm. The backward view for the recursion provides

$$Y(n) = Y^1(I(n) - 1) + Y^2(n - I(n)) + n - 1,$$

$Y^1, Y^2, I(n)$ independent, $Y^1 \overset{\mathcal{D}}{=} Y^2$, $I(n)$ uniformly on $\{1, 2, \ldots, n\}$ (=the rank of U_1 under $U_1, \ldots, U_n$).

Then $X(n) := \frac{Y(n) - EY(n)}{n+1}$ satisfies

$$X(n) \overset{\mathcal{D}}{=} \frac{I(n)}{n+1} X^1(I(n)) + \frac{n - I(n)}{n+1} X^2(n + 1 - I(n)) + C(I(n)).$$

Since $\frac{I(n)}{n+1}$ converges to a uniformly distributed rv U and $C(I(n)) \to_n C(U)$ we expect in the limit the Quicksort recursion

$$X \overset{\mathcal{D}}{=} UX^1 + (1 - U)X^2 + C(U).$$

There is a unique solution in $L^p,\ p > 1$, the Quicksort distribution.

Rem: Régnier [34] found that $(X(n))_n$ itself is an L^2-martingale and $X(n)$ converges a.e. to a non degenerate limit. This is due to extremely nice properties of the binary search trees.

13.6 Process Valued WBP and SFE Solutions

We consider now processes $X = X(t)_{t \in [0,1]}$ in Skorodhod space of càdlàg functions D as solution of some SFE. Many of the examples are of the form, take an interval splitting (WBP) $I_v, v \in V$, of the unit interval and another WBP $(V, L, G, *)$. Consider the stochastic process $Z_n = \sum_{|v|=n} 1\!\!1_{I_v} L_v$ in the limit. Examples are the Mandelbrot Cascades or the Find process.

Here we take some $G \subset D^D$ with composition $x * y = y \circ x$. Since the intervals are lexicographically ordered, we may interpret t as time. A typical example is $(a, b) \in G$ given by a space transformation $a \in D$ and a time transformation b. The formal description is $(a, b) * (c, d) = (ac \circ b, b \circ d)$ and $(a, b) *_r f = af \circ b$ for $f \in D$. Here are some limiting examples for Z_n and $R_n = \sum_{|v|<n} L_v *_r C_v$ satisfying a SFE.

Ex: **Brownian motion**

$$X \overset{\mathrm{d}}{=} (1\!\!1_{t<1/2} \frac{X_1(2t)}{\sqrt{2}} + 1_{t \geq 1/2}(X_1(1) + \frac{1}{\sqrt{2}} X_2(2t-1)))_t.$$

Ex: **Find** Find is an algorithm to find the l-th largest within n elements [20].

$$X \overset{\mathcal{D}}{=} (1\!\!1_{t<U} U X_1(\frac{t}{U}) + 1\!\!1_{U \leq t}(1-U) X_2(\frac{t-U}{1-U}) + 1)_t,$$

U uniformly distributed on $[0, 1]$

Ex: **Quicksort process** [38]

$$X \overset{\mathcal{D}}{=} (1\!\!1_{U>t} U X_1(1 \wedge \frac{t}{U}) + 1\!\!1_{U \leq t}(1-U) X_2(\frac{t-U}{1-U}) + C(U,t))_t,$$

$$\begin{aligned} C(x,t) &= C(x) + 2 1\!\!1_{x \geq t}((1-t)\ln(1-t) + (1-x)\ln(1-x) + 1) \\ &\quad -(x-t)\ln(x-t), \\ C(x) &= 1 + 2x \ln x + 2(1-x)\ln(1-x). \end{aligned}$$

The Quicksort process is the unique solution of the above SFE with centered rvs $X(t)$ having finite second moments for all t.

The splitting is obvious in all examples. The Brownian example is non endogenous, since considering only $t = 1$ we are back to the corresponding SFE for the normal distribution. The Find example is the limit of the rvs $R_n = \sum_{|v|<n} 1\!\!1_{I_v} |I_v|$ and that solution is endogenous. The Quicksort process example looks like a splitting with overlapping intervals within the same generation. Via a reformulation [38] this example is of the form $R_n = \sum_{|v|<n} 1\!\!1_{I_v} L_v *_r C_v$ for suitable random variables C_v with values in D, a non overlapping splitting I_v of the unit interval by a uniformly distributed rvs and $L_{v,vi}$ is the obvious space-time transformation. The solution of the SFE is endogenous.

In general Knof (unpublished) considered in his Ph.D. overlapping splittings and SFEs on D with equality for finite marginals only. Sulzbach [31] took an approach via a functional contraction method using the Zolotarev metric.

Discrete Quicksort Process There is also a discrete Quicksort process version.

Conrado Martínez introduced Partial Quicksort: For a sequence of n different reals find the smallest l reals in natural order. The procedure is: Recall Quicksort always for the left most list with 2 or more elements. Publish first the smallest number then second smallest and so on.

Observation: In order to find the l smallest numbers in order the algorithm does only necessary comparisons.

Therefore consider l as time. Let $Y(n,l)$ be the number of comparisons for Partial Quicksort. Then consider the normalized rv

$$X(n, \frac{l}{n}) = \frac{Y(n,l) - EY(n,l)}{n+1}$$

extended to D. They satisfying some 'noisy' fixed point equation on D, compare to the discrete Quicksort. The best available and possible result is:

Theorem 13.3 (Roesler) *Let U_i, $i \in \mathbb{N}$, be independent uniformly distributed rvs. Then $X(n,\cdot)$ converges almost surely to a specific version of the Quicksort process in Skorodhod metric on D.*

References

1. Aldous, D., Bandyopadhyay, A.: A survey of max-type recursive distributional equations. Ann. Appl. Probab. **15**, 1047–1110 (2005)
2. Alsmeyer, G.: The smoothing transform: a review of contraction results. In: Alsmeyer, G., Löwe, M. (eds.) Random Matrices and Iterated Random Functions. Springer Proceedings in Mathematics & Statistics, vol. 53, pp. 189–228. Springer, Berlin (2013)
3. Alsmeyer, G., Kuhlbusch, D.: Double martingale structure and existence of f-moments for weighted branching processes. Preprint no. 08/05-S. Universität Münster, Germany (2006)
4. Alsmeyer, G., Meiners, M.: Fixed point of inhomogeneous smoothing transforms. arXiv: 1007.4509v2 (2011)
5. Alsmeyer, G., Meiners, M.: Fixed points of the smoothing transform: two-sided solutions. Probab. Theory Relat. Fields **155**, 165–199 (2013)
6. Alsmeyer, A., Roesler, U.: A stochastic fixed point equation related to weighted branching with deterministic weights. Electron J. Probab. **11**, 27–56 (2005)
7. Alsmeyer, G., Roesler, U.: A stochastic fixed point equation for weighted minima and maxima. Ann. Henri. Poincaré **44**, 89–103 (2008)
8. Alsmeyer, G., Biggins, J.D., Meiners, M.: The functional equation of the smoothing transform. Ann. Probab. **40**, 2069–2105 (2012)
9. Athreya, K.B., Ney, P.E.: Branching Processes. Die Grundlehren der mathematischen Wissenschaften, vol. 196. Springer, Berlin (1972)
10. Attal, S.: Markov chains and dynamical systems: the open system point of view. arXiv:1010.2894 (2010)

11. Barnsley, M.F.: Fractals Everywhere, 2nd edn. AP Professional, New York (1993). ISBN-10: 0120790610, ISBN-13: 978-0120790616
12. Barnsley, M.F.: SuperFractals, 2nd edn. Cambridge University Press, Cambridge (2006). ISBN: 978-0521844932
13. Barral, J., Mandelbrot, B.: Fractional multiplicative processes. Ann. Inst. Henri Poincaré Probab. Stat. **45**, 1116–1129 (2009)
14. Biggins, J.D.: Martingale convergence in the branching random walk. J. Appl. Prob. **14**, 25–37 (1977)
15. Biggins, J.D., Kyprianou, A.E.: Measure change in multitype branching. Adv. Appl. Probab. **36**, 544–481 (2004)
16. Burton, R., Roesler, U.: An L_2 convergence theorem for random affine mappings. J. App. Probab. **32**, 183–192 (1995)
17. Durrett, R., Liggett, T.M.: Fixed points of the smoothing transform. Z. Wahrscheinlichkeitstheorie verw. Gebiete **64**, 275–301 (1983)
18. Falconer, K.: Fractal Geometry: Mathematical Foundations and Applications, 2nd edn. Wiley, New York (2003)
19. Fill, J.A., Janson, S.: A characterization of the set of fixed points of the Quicksort transformation. Electron. Commun. Prob. **5**, 77–84 (2000)
20. Grübel, R., Roesler, U. Asymptotic distribution theory for Hoare's selection algorithm. Adv. Appl. Probab. **28**, 252–269 (1996)
21. Holley, R., Liggett, T.M.: Generalized potlatch and smoothing processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete **55**, 165–195 (1981)
22. Iksanov, A., Meiners, M.: Fixed points of multivariate smoothing transforms with scalar weights. arXiv 1402.4147v1 (2014)
23. Kahane, J-P., Peyriére, J.: Sur Certaines Martingals de Benoît Mandelbrot. Adv. Math. **22**, 131–145 (1976)
24. Kellerer, H.G.: Ergodic behaviour of affine recursions I-III Preprints (1992)
25. Liu, Q.: Fixed points of a generalized smoothing transformation and applications to the branching random walk. Adv. Appl. Probab. **30**, 85–112 (1998)
26. Liu, Q.: On generalized multiplicative cascades. Stochastic Process. Appl. **86**, 263–286 (2000)
27. Lyons, R.: A simple path to Biggins's martingale convergence for branching random walk. In: Athreya, K.B., Jagers, P. (eds.) Classical and Modern Branching Processes. IMA Volumes in Mathematics and its Applications, vol. **84**, pp. 217–221. Springer, Berlin (1997)
28. Neininger, R.: Probabilistic analysis of algorithms, stochastic fixed point equations and ideal metrics. Lecture Notes: Summer School on Probabilistic Methods in Combinatorics Graz, July 17–19 (2006)
29. Neininger, R., Rüschendorf, L.: General limit theorem for recursive algorithms and combinatorial structures. Ann. Appl. Probab. **14**, 378–418 (2004)
30. Neininger, R., Rüschendorf, L.: Analysis of algorithms by the contraction method: additive and max-recursive sequences. In: Interacting Stochastic Systems, pp. 435–450. Springer, Berlin (2005)
31. Neininger, R., Sulzbach, H.: On a functional contraction method. Ann. Probab. **43**, 1777–1822 (2015)
32. Pachter, L., Sturmfels, B.: Algebraic Statistics for Computational Biology. Cambridge University Press, Cambridge (2005). ISBN-13: 978-0521857000
33. Rachev, S.T., Rüschendorf, L.: Probability metrics and recursive algorithms. Adv. Appl. Probab. **27**, 770–799 (1995)
34. Régnier, M.: A limiting distribution for quicksort. RAIRO Theor. Inform. Appl. **23**, 335–343 (1989)
35. Roesler, U.: A limit theorem for "Quicksort". RAIRO Theor. Inform. Appl. **25**, 85–100 (1991)
36. Roesler, U.: A fixed point theorem for distributions. Stoch. Process. Appl. **42**, 195–214 (1992)
37. Roesler, U.: The weighted branching process. In: Dynamics of Complex and Irregular Systems (Bielefeld, 1991): Bielefeld Encounters in Mathematics and Physics VIII, pp. 154–165. World Scientific Publishing, River Edge, NJ (1993)

38. Roesler, U.: Almost sure convergence to the Quicksort process. Preprint (2015)
39. Roesler, U., Rüschendorf, L.: The contraction method for recursive algorithms. Algorithmica **29**, 3–33 (2001)
40. Rüschendorf, L.: On stochastic recursive equations of sum and max type. J. Appl. Probab. **43**, 687–703 (2006)
41. Waymire, E., Williams, S.: Markov cascades. In: Athreya, K.B., Jagers, P. (eds.) Classical and Modern Branching Processes, IMA Volumes in Mathematics and its Applications. vol. 84, pp. 305–321. Springer, Berlin (1997)

Chapter 14
A Special Family of Galton-Watson Processes with Explosions

Serik Sagitov and Alexey Lindo

Mathematics Subject Classification (2000): 60J80

14.1 Introduction

Consider a Galton-Watson process $(Z_n)_{n\geq 0}$ with $Z_0 = 1$ and the offspring number distribution

$$p_k = P(Z_1 = k), \quad k \geq 0.$$

The properties of this branching process are studied in terms of the probability generating function

$$f(s) = p_0 + p_1 s + p_2 s^2 + \dots,$$

where it is usual to assume that $f(1) = 1$, however, in this paper we allow for $f(1) < 1$ so that a given particle may explode with probability $p_\infty = 1 - f(1)$. The probability generating function $f_n(s) = E(s^{Z_n})$ of the size of the n-th generation is given by the n-fold iteration of $f(s)$

$$f_0(s) = s, \quad f_n(s) = f(f_{n-1}(s)), \quad n \geq 1,$$

S. Sagitov (✉) • A. Lindo
Chalmers University and University of Gothenburg, SE-412 96 Gothenburg, Sweden
e-mail: serik@chalmers.se; lindo@chalmers.se; alexey.lindo@gmail.com

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_14

and therefore it is desirable to have a range of probability generating functions f whose iterations can be computed explicitly.

The best known case of explicit calculations is the family of linear-fractional Galton-Watson processes with

$$f(s) = p_0 + (1 - p_0)\frac{ps}{1 - (1 - p)s}, \quad s \in [0, (1 - p)^{-1}),$$

representing the family of modified geometric distributions

$$p_k = (1 - p_0)(1 - p)^{k-1}p, \quad k \geq 1,$$

fully characterized by just two parameters: $p_0 \in [0, 1)$ and $p \in (0, 1]$. In Sect. 14.2 for each $\theta \in [-1, 1]$ we introduce a family $\mathcal{G}_\theta$ of functions with explicit iterations containing the linear-fractional family as a particular case. In Sect. 14.3 we demonstrate that all $f \in \mathcal{G}_\theta$ are probability generating functions with $f(1) \leq 1$. A Galton-Watson processes with the reproduction law whose probability generating function belongs to $\mathcal{G}_\theta$ will be called a theta-branching process.

The basic properties of the theta-branching processes are summarized in Sect. 14.4, where it is shown that this family is wide enough to include the cases of infinite variance, infinite mean, and even non-regular branching processes with explosive particles.

Recall that the basic classification of the Galton-Watson processes refers to the mean offspring number $m = EZ_1$. Let $q \in [0, 1]$ be the smallest non-negative root of the equation $f(x) = x$ and denote by

$$T_0 = \inf\{n : Z_n = 0\}$$

the extinction time of the branching process. Then $q = P(T_0 < \infty)$ gives the probability of ultimate extinction. For $m \leq 1$ and $p_1 < 1$, the extinction probability is $q = 1$, while in the supercritical case $m > 1$, we have $q < 1$.

If $f(1) < 1$, then the Galton-Watson process is a Markov chain with two absorption states $\{0\}$ and $\{\infty\}$. In this case the branching process either goes extinct at time T_0 or explodes at the time

$$T_1 = \inf\{n : Z_n = \infty\},$$

with

$$P(T_1 \leq n) = 1 - f_n(1), \quad P(T_1 < \infty) = 1 - q,$$

where the latter equality is due to $f_n(1) \to q$. In Sect. 14.5, using explicit formulas for $f_n(s)$ we compute the distribution of the absorption time

$$T = T_0 \wedge T_1.$$

Note that in the regular case, we have $P(T_1 = \infty) = 1$ and therefore, $T \equiv T_0$. Observe also that the case $f(1) < 1$ has other, biologically more relevant interpretations. For example in the multitype setting, T_1 can be viewed as the time of the first mutation event, see [7].

Also in Sect. 14.5 we consider a situation when the explosion of a single particle has a small probability, so that T_1 takes large values in explosion scenarios. We show that in such a case the time to explosion can be asymptotically characterized with help of a Gumbel distribution. In Sect. 14.6 we study the Q-processes for the theta-branching processes extending the classical definition to the non-regular case. Our explicit calculations demonstrate that in the non-regular case the behavior of a branching process is more similar to that of the subcritical rather than supercritical regular case. Using these results on the Q-processes we derive the conditional limits of the theta-branching processes conditioned on non-absorption.

A remarkable property of the linear-fractional Galton-Watson processes is that they can be embedded into the linear birth-death processes. In Sect. 14.7 we establish embeddability of theta-branching processes.

14.2 Probability Generating Functions for Theta-Branching Processes

Using an alternative parametrization for the linear-fractional probability generating functions, we obtain

$$\frac{1}{1-f(s)} = \frac{a}{1-s} + c, \quad s \in [0,1), \tag{14.1}$$

where

$$a = \frac{p}{1-p_0}, \quad c = \frac{1-p}{1-p_0}.$$

This observation immediately implies that the n-fold iteration f_n of the linear-fractional f is also linear-fractional

$$\frac{1}{1-f_n(s)} = \frac{a^n}{1-s} + c(1 + a + \ldots + a^{n-1}).$$

The key idea of this paper is to expand the family (14.1) by

$$(A-f(s))^{-\theta} = a(A-s)^{-\theta} + c, \quad s \in [0,A), \tag{14.2}$$

with the help of two extra parameters (A, θ) which are invariant under iterations.

Definition 14.1 Let $\theta \in (-1, 0) \cup (0, 1]$. We say that a probability generating function f belongs to the family $\mathcal{G}_\theta$ if

$$f(s) = A - [a(A-s)^{-\theta} + c]^{-1/\theta}, \qquad 0 \le s < A,$$

where one of the following three options holds

$$\begin{array}{lll} (i)\ \ a \ge 1, & c > 0, & \theta \in (0,1],\ A = 1, \\ (ii)\ \ a \in (0,1), & c = (1-a)(1-q)^{-\theta}, & q \in [0,1),\ A = 1, \\ (iii)\ \ a \in (0,1), & c = (1-a)(A-q)^{-\theta}, & q \in [0,1],\ A > 1. \end{array}$$

Definition 14.1 can be extended to the case $\theta = 0$ by the following continuity argument: for $a \in (0, 1)$

$$A - [a(A-s)^{-\theta} + (1-a)(A-q)^{-\theta}]^{-1/\theta} \to A - (A-q)^{1-a}(A-s)^a, \quad \theta \to 0.$$

Definition 14.2 We say a probability generating function f belongs to

- the family $\mathcal{G}_0$ if for some $a \in (0, 1)$,

  $$f(s) = A - (A-q)^{1-a}(A-s)^a, \qquad 0 \le s < A,$$

  where either $A = 1$, $q \in [0, 1)$, or $A > 1$, $q \in [0, 1]$,
- the family $f \in \mathcal{G}_{-1}$ if for some $q \in [0, 1]$ and $a \in (0, 1)$,

  $$f(s) = as + (1-a)q, \qquad 0 \le s < \infty.$$

Definition 14.3 A Galton-Watson process with the reproduction law whose probability generating function $f \in \mathcal{G}_\theta$, $\theta \in [-1, 1]$, will be called a theta-branching process.

It is straightforward to see, cf. Sect. 14.4, that each of the families $\mathcal{G}_\theta$ is invariant under iterations: if $f \in \mathcal{G}_\theta$, then $f_n \in \mathcal{G}_\theta$ for all $n \ge 1$. The fact, that the functional families in Definitions 14.1 and 14.2 are indeed consist of probability generating functions with $f(1) \le 1$, is verified in Sect. 14.3.

Parts of the $\mathcal{G}_\theta$ families were mentioned earlier in the literature as examples of probability generating functions with explicit iterations. Clearly, $\mathcal{G}_1 \cup \mathcal{G}_{-1}$ is the family of linear-fractional probability generating functions. Examples in [10] leads to the case $A = 1$ and $\theta \in [0, 1)$, which was later given among other examples in Chap. 1.8 of [8]. The case $A = 1$ and $\theta \in (0, 1)$ was later studied in [9]. A special pdf with $\theta = -1/2$,

$$f(s) = 1 - (a\sqrt{1-s} + 1 - a)^2, \quad a \in (0, 1),$$

can be found in [2] on page 112, as an example of non-regular Galton-Watson processes.

Notice that there is a version of linear-fractional Galton-Watson processes with countably many types of particles, see [5]. It is an open problem to expand the theta-branching processes with $\theta \in (-1, 1)$ to the multitype setting.

14.3 Monotonicity Properties

It is straightforward to see that each $f \in \mathcal{G}_0$ is a probability generating function with

$$f'(s) = (A-q)^{1-a} a (A-s)^{a-1},$$
$$f^{(n)}(s) = (A-q)^{1-a} a(1-a)\dots(n-1-a)(A-s)^{a-n},\ n \geq 2,$$

and

$$p_0 = A - (A-q)^{1-a} A^a,$$
$$p_1 = (A-q)^{1-a} a A^{a-1},$$
$$p_n = p_{n-1} \frac{n-a-1}{nA}, \quad n \geq 2.$$

Therefore, $(p_n)_{n \geq 1}$ are monotonely decreasing with

$$p_n = aA^a (A-q)^{1-a} A^{-n} \prod_{k=2}^{n} \Big(1 - \frac{1+a}{k}\Big), \quad n \geq 2,$$

so that $p_n \sim \text{const} \cdot A^{-n} n^{-1-a}$ as $n \to \infty$.

Proposition 14.1 *Let* $\theta \in (-1, 0) \cup (0, 1)$ *and* $f \in \mathcal{G}_\theta$. *Then* f *is a probability generating function with* $f(1) \leq 1$ *such that*

$$p_0 = A - (aA^{-\theta} + c)^{-1/\theta},$$
$$p_1 = a(a + cA^\theta)^{-1-1/\theta},$$

and for $n \geq 2$,

$$p_n = \frac{aA^{-n+1}}{(a + cA^\theta)^{\frac{1+\theta}{\theta}} n!} \cdot \sum_{i=1}^{n-1} \Big(\frac{cA^\theta}{a + cA^\theta}\Big)^i B_{i,n},$$

where all $B_{i,n} = B_{i,n}(\theta)$ *are non-negative and, for* $n \geq 2$*, satisfy the recursion*

$$B_{i,n} = (n-2-i\theta)B_{i,n-1} + (1+i\theta)B_{i-1,n-1},\ i = 1,\dots,n-1,$$

with $B_{0,n} = B_{n,n} = 0$ *for* $n \geq 1$*, and* $B_{1,2} = 1+\theta$*.*

Proof In terms of

$$\phi(s) := \frac{A - f(s)}{A - s} = [a + c(A-s)^{\theta}]^{-1/\theta}, \quad \phi'(s) = c(A-s)^{\theta-1}\phi(s)^{1+\theta},$$

we have

$$\begin{aligned}
f'(s) &= a\phi(s)^{1+\theta},\\
f''(s) &= (1+\theta)ac(A-s)^{\theta-1}\phi(s)^{1+2\theta},\\
f'''(s) &= (1+\theta)(1-\theta)ac(A-s)^{\theta-2}\phi(s)^{1+2\theta}\\
&\quad + (1+\theta)(1+2\theta)ac^2(A-s)^{2\theta-2}\phi(s)^{1+3\theta}.
\end{aligned}$$

and more generally,

$$f^{(n)}(s) = \sum_{i=1}^{n-1} B_{i,n}ac^i(A-s)^{i\theta-n+1}\phi(s)^{1+(i+1)\theta}, \quad n \geq 2,$$

where $B_{i,n}$ are defined in the statement. To finish the proof it remains to apply the equality $p_n = f^{(n)}(0)/n!$

In the linear-fractional case we have $p_k \geq p_{k+1}$ for all $k \geq 1$. The next extension of this monotonicity property was first established in [9].

Corollary 14.1 *Let* $\theta \in (0,1)$ *and* $f \in \mathcal{G}_\theta$ *with* $A = 1$*. Then* $p_k \geq p_{k+1}$ *for all* $k \geq 1$*.*

Proof Put

$$g(s) = (s-1)f(s) = -p_0 + \sum_{k=1}^{\infty}(p_{k-1} - p_k)s^k$$

From

$$\begin{aligned}
g(s) &= s - 1 + (1-s)^2[a + c(1-s)^{\theta}]^{-1/\theta},\\
g'(s) &= 1 + c(1-s)^{\theta+1}[a + c(1-s)^{\theta}]^{-1-1/\theta} - 2(1-s)[a + c(1-s)^{\theta}]^{-1/\theta}\\
&= c(1-f(s))^{1+\theta} + 2f(s) - 1,\\
g''(s) &= (2 - c(1+\theta)(1-f(s))^{\theta})f'(s),
\end{aligned}$$

we see that $g''(s) \geq 0$, since

$$G(s) := 2 - c(1+\theta)(1-f(s))^{\theta} \geq 2 - c(1+\theta)(1-p_0)^{\theta} = 2 - \frac{c(1+\theta)}{a+c} > 0.$$

Furthermore,

$$G'(s) = c\theta(1+\theta)(1-f(s))^{\theta-1}f'(s)$$

is absolutely monotone (as a product of two absolutely monotone functions), implying that $g''(s)$ is absolutely monotone, so that

$$k(k-1)(p_{k-1} - p_k) \geq 0, \quad k \geq 2.$$

14.4 Basic Properties of $f \in \mathcal{G}_\theta$

In this section we distinguish among nine cases inside the collection of families $\{\mathcal{G}_\theta\}_{-1 \leq \theta \leq 1}$ and summarize the following basic formulas: $f_n(s), f(1), f'(1), f''(1)$. In all cases, except Case 1, we have $a = f'(q)$. The following definition, cf [3], explains an intimate relationship between the Cases 3–5 with $A = 1$ and the Cases 7–9 with $A > 1$.

Definition 14.4 Let $A > 1$ and a probability generating function f be such that $f(A) \leq A$. We call

$$\hat{f}(s) := \frac{f(sA)}{A} = \sum_{k=0}^{\infty} p_k A^{k-1} s^k$$

the dual generating function for f and denote $\hat{q} = qA^{-1}$, so that $\hat{f}(\hat{q}) = \hat{q}$. Clearly, $\hat{f}'(\hat{q}) = f'(q)$.

Case 1: $\theta \in (0,1]$, $a \in (1,\infty)$,

$$f_n(s) = 1 - [a^n(1-s)^{-\theta} + (a^n - 1)d]^{-1/\theta}, \quad d \in (0,\infty).$$

The corresponding theta-branching process is subcritical with $m = a^{-1/\theta}$. If $\theta \in (0,1)$, then $f''(1) = \infty$ and for $\theta = 1$ we have $f''(1) = 2(a-1)a^{-2}d$.

Case 2: $\theta \in (0,1]$, $a = 1$,

$$f_n(s) = 1 - [(1-s)^{-\theta} + nc]^{-1/\theta}, \quad c \in (0,\infty).$$

The corresponding theta-branching process is critical with either finite or infinite variance. If $\theta \in (0,1)$, then $f''(1) = \infty$ and for $\theta = 1$ we have $f''(1) = 2c$. This is the only critical case in the whole family of theta-branching process.

Case 3: $\theta \in (0,1]$, $a \in (0,1)$,

$$f_n(s) = 1 - \left[a^n(1-s)^{-\theta} + (1-a^n)(1-q)^{-\theta}\right]^{-1/\theta}, \quad q \in [0,1).$$

The corresponding theta-branching process is supercritical with $m = a^{-1/\theta}$. If $\theta \in (0,1)$, then $f''(1) = \infty$, and for $\theta = 1$ we have $f''(1) = 2a^{-2}(1-a)(1-q)^{-1}$.
Case 4: $\theta = 0$, $a \in (0,1)$,

$$f_n(s) = 1 - (1-q)^{1-a^n}(1-s)^{a^n}, \quad q \in [0,1).$$

The theta-branching process is regular supercritical with infinite mean.
Case 5: $\theta \in (-1,0)$, $a \in (0,1)$,

$$f_n(s) = 1 - \left[a^n(1-s)^{|\theta|} + (1-a^n)(1-q)^{|\theta|}\right]^{1/|\theta|}, \quad q \in [0,1).$$

The theta-branching process is non-regular with a positive

$$1 - f(1) = (1-a)^{1/|\theta|}(1-q)$$

and infinite $f'(1)$.
Case 6: $\theta = -1$, $a \in (0,1)$,

$$f_n(s) = a^n s + (1-a^n)q, \quad q \in [0,1].$$

If $q = 1$, then the theta-branching process becomes a pure death process with mean $m = a$ and $f''(1) = 0$. If $q < 1$, then the θ-process is non-regular with a positive

$$1 - f(1) = (1-a)(1-q),$$

$f'(1) = a$ and $f''(1) = 0$.
Case 7: $\theta \in (0,1]$, $a \in (0,1)$, $A > 1$,

$$f_n(s) = A - [a^n(A-s)^{-\theta} + (1-a^n)(A-q)^{-\theta}]^{-1/\theta}, \quad q \in [0,1].$$

If $q = 1$, then the corresponding theta-branching process is subcritical with the offspring mean $m = a$ and

$$f''(1) = (1+\theta)a(1-a)(A-1)^{-1}.$$

If $q \in [0,1)$, the theta-branching process is non-regular with a positive

$$1 - f(1) = (A-1)([a + (1-a)(A-q)^{-\theta}(A-1)^{\theta}]^{-1/\theta} - 1),$$

and

$$f'(1) = a[a + (1-a)(A-q)^{-\theta}(A-1)^{\theta}]^{-1/\theta-1},$$

$$f''(1) = (1+\theta)a(1-a)(A-q)^{-\theta}(A-1)^{\theta-1}[a+(1-a)(A-q)^{-\theta}(A-1)^{\theta}]^{-1/\theta-2}.$$

We have $f(A) = A$, and the dual generating function has the form of the Case 3:

$$\hat{f}(s) = 1 - [a(1-s)^{-\theta} + (1-a)(1-\hat{q})^{-\theta}]^{-1/\theta}.$$

Case 8: $\theta = 0, a \in (0,1), A > 1,$

$$f_n(s) = A - (A-q)^{1-a^n}(A-s)^{a^n}, \quad q \in [0,1].$$

If $q = 1$, the theta-branching process is subcritical with the offspring mean $m = a$ and

$$f''(1) = a(1-a)(A-1)^{-1}.$$

If $q \in [0,1)$, the theta-branching process is non-regular with a positive

$$1 - f(1) = (A-q)^{1-a}(A-1)^{a} - (A-1),$$

and

$$f'(1) = a(A-q)^{1-a}(A-1)^{a-1},$$

$$f''(1) = a(1-a)(A-q)^{1-a}(A-1)^{a-2}.$$

We have $f(A) = A$, and the dual generating function belongs to the Case 4:

$$\hat{f}(s) = 1 - (1-\hat{q})^{1-a}(1-s)^{a}.$$

Case 9: $\theta \in (-1,0), a \in (0,1), A > 1,$

$$f_n(s) = A - \left[a^n(A-s)^{|\theta|} + (1-a^n)(A-q)^{|\theta|}\right]^{1/|\theta|}, \quad q \in [0,1].$$

If $q = 1$, then the theta-branching process is subcritical with the offspring mean $m = a$ and

$$f''(1) = (1-|\theta|)a(1-a)(A-1)^{-1}.$$

If $q \in [0,1)$, the theta-branching process is non-regular with a positive

$$1 - f(1) = [a(A-1)^{|\theta|} + (1-a)(A-q)^{|\theta|}]^{1/|\theta|} - (A-1),$$

and

$$\begin{aligned} f'(1) &= a[a + (1-a)(A-q)^{|\theta|}(A-1)^{-|\theta|}]^{1/|\theta|-1} \in (0,1), \\ f''(1) &= (1-|\theta|)a(1-a)(A-q)^{|\theta|}(A-1)^{-|\theta|-1} \\ &\quad \cdot [a + (1-a)(A-q)^{|\theta|}(A-s)^{-|\theta|}]^{1/|\theta|-2}. \end{aligned}$$

With

$$f(A) = A - (1-a)^{1/|\theta|}(A-q) \in (q, A),$$

the dual generating function takes the form of the Case 5:

$$\hat{f}(s) = 1 - [a(1-s)^{|\theta|} + (1-a)(1-\hat{q})^{|\theta|}]^{1/|\theta|}.$$

14.5 Extinction and Explosion Times

Recall that $T = T_0 \wedge T_1$, and in the regular case $T = T_0$. In the non-regular case, when $f(1) < 1$, from

$$\begin{aligned} P(n < T_0 < \infty) &= q - f_n(0), \\ P(n < T_1 < \infty) &= f_n(1) - q, \end{aligned}$$

we obtain

$$P(n < T < \infty) = f_n(1) - f_n(0).$$

For our special family of branching processes we compute explicitly the distribution functions of the times T_0, T_1, T.

Cases 1–4. In these regular cases we are interested only in the extinction time:

$$P(n < T_0 < \infty) = \begin{cases} a^{-n/\theta}[1 + d - da^{-n}]^{-1/\theta}, & \text{Case 1}, \\ (1+cn)^{-1/\theta}, & \text{Case 2}, \\ (1-q)\big([1 - a^n(1-(1-q)^{\theta})]^{-1/\theta} - 1\big), & \text{Case 3}, \\ (1-q)[(1-q)^{-a^n} - 1], & \text{Case 4}. \end{cases}$$

Cases 5, 7, 9. In these cases

$$\begin{aligned} P(n < T_0 < \infty) &= (A-q)([1 - a^n(1-(A-q)^{\theta}A^{-\theta})]^{-1/\theta} - 1), \\ P(n < T_1 < \infty) &= (A-q)\big(1 - [1 - a^n(1-(A-q)^{\theta}(A-1)^{-\theta})]^{-1/\theta}\big), \end{aligned}$$

$$P(n < T < \infty) = (A-q)\Big\{[1 - a^n(1-(A-q)^\theta A^{-\theta})]^{-1/\theta} \\ - [1 - a^n(1-(A-q)^\theta (A-1)^{-\theta})]^{-1/\theta}\Big\}.$$

Case 6. In this trivial case

$$P(n < T_0 < \infty) = a^n q, \quad P(n<T_1 < \infty) = a^n(1-q), \quad P(n < T < \infty) = a^n.$$

and for $q \in (0,1)$,

$$E(T_0|T_0 < \infty) = E(T_1|T_1 < \infty) = E(T) = \frac{1}{1-a}.$$

Case 8. In this case

$$\begin{aligned} P(n < T_0 < \infty) &= (A-q)[(A-q)^{-a^n} A^{a^n} - 1], \\ P(n < T_1 < \infty) &= (A-q)[1 - (A-q)^{-a^n}(A-1)^{a^n}], \\ P(n < T < \infty) &= (A-q)^{1-a^n}[A^{a^n} - (A-1)^{a^n}]. \end{aligned}$$

Theorem 14.1 *Consider a theta-branching process with* $\theta \in (-1, 0]$ *and* $A \geq 1$. *Let* $\theta \to 0$ *and* $A \to 1$ *in such a way that*

$$|\theta| \cdot \log \frac{1}{A-1} \to r, \quad r \in [0, \infty].$$

Then for any fixed $a \in (0,1)$, $q \in [0,1)$, *and* $y \in (-\infty, \infty)$,

$$\lim_{\epsilon \to 0} P(T_1 - \log_a \epsilon \leq y | T_1 < \infty) = e^{-wa^y},$$

where

$$\epsilon = \begin{cases} |\theta|, & r \in (0,\infty], \\ (\log \frac{1}{A-1})^{-1}, & r = 0, \end{cases} \qquad w = \begin{cases} 1, & r \in \{0\} \cup \{\infty\}, \\ 1 - e^{-r}, & r \in (0, \infty). \end{cases}$$

The limit is a Gumbel distribution with mean $\frac{\log w - \gamma}{\log a}$, *where* γ *is the Euler–Mascheroni constant.*

Proof In view of

$$P(T_1 \leq n | T_1 < \infty) = \frac{A-q}{1-q}\big[1 - a^n(1-(A-1)^{|\theta|}(A-q)^{-|\theta|})\big]^{1/|\theta|} - \frac{A-1}{1-q},$$

it suffices to verify that

$$\left[1-\epsilon a^y(1-(A-1)^{|\theta|})\right]^{1/|\theta|}\to e^{-wa^y}.$$

Indeed, if $r=\infty$, then $(A-1)^{|\theta|}\to 0$, and

$$\left[1-|\theta|a^y(1-(A-1)^{|\theta|})\right]^{1/|\theta|}\to e^{-a^y}.$$

If $r\in(0,\infty)$, then $(A-1)^{|\theta|}\to e^{-r}$, and

$$\left[1-|\theta|a^y(1-(A-1)^{|\theta|})\right]^{1/|\theta|}\to e^{-a^y(1-e^{-r})}.$$

Finally, if $r=0$, then

$$1-(A-1)^{|\theta|}\sim|\theta|/\epsilon,$$

and therefore

$$\left[1-\epsilon a^y(1-(A-1)^{|\theta|})\right]^{1/|\theta|}\to e^{-a^y}.$$

Corollary 14.2 *If $A=1$ and $\theta\in(-1,0)$, then for any fixed $a\in(0,1)$ and $q\in[0,1)$,*

$$\lim_{\theta\to 0}P(T_1-\log_a|\theta|\le y|T_1<\infty)=e^{-a^y},\quad y\in(-\infty,\infty),$$

If $\theta=0$ and $A=1+e^{-1/\epsilon}$, $\epsilon>0$, then for any fixed $a\in(0,1)$ and $q\in[0,1)$,

$$\lim_{\epsilon\to 0}P(T_1-\log_a\epsilon\le y|T_1<\infty)=e^{-a^y},\quad y\in(-\infty,\infty).$$

14.6 The *Q*-Process

As explained in Chap. I.14, [1], for a regular Galton-Watson process with transition probabilities $P_n(i,j)$, one can define another Markov chain with transition probabilities

$$Q_n(i,j):=\frac{jq^{j-i}P_n(i,j)}{\gamma^n i},\quad i\ge 1,\ ,j\ge 1,$$

where $\gamma=f'(q)$. The new chain is called the Q-process, and from

$$\sum_{j\ge 1}Q_n(i,j)s^j=\frac{s}{\gamma^n iq^i}\frac{d}{ds}(f_n^i(sq))=s\cdot\frac{f_n'(sq)}{f_n'(q)}\cdot\left(\frac{f_n(sq)}{q}\right)^{i-1}$$

we see that the Q-process is a Galton-Watson process with the dual reproduction $\frac{f(sq)}{q}$ and an eternal particle generating a random number κ of ordinary particles with $E(s^\kappa) = \frac{f'(sq)}{f'(q)}$, see [3]. The Q-process in the regular case is interpreted in [1] as the original branching process "conditioned on not being extinct in the distant future and on being extinct in the even more distant future".

Exactly the same definition of the Q-process makes sense in the non-regular case, only now the last interpretation should be based on the absorption time T rather than on the extinction time T_0. Indeed, writing $P_j(\cdot) = P(\cdot|Z_0 = j)$ we get for $j \geq 1$,

$$P_j(T > n) = f_n^j(1) - f_n^j(0),$$

and therefore,

$$P_i(Z_1 = j_1, \dots, Z_n = j_n | T > n + k) = P_i(Z_1 = j_1, \dots, Z_n = j_n) \frac{f_k^{j_n}(1) - f_k^{j_n}(0)}{f_{n+k}^i(1) - f_{n+k}^i(0)}.$$

In the non-regular case, as $k \to \infty$ we have $f_k(0) \to q$ and $f_k(1) \to q$. Thus, repeating the key argument of Chap. I.14, [1] for the derivation of the Q-process,

$$P_i(Z_1 = j_1, \dots, Z_n = j_n | T > n + k) \to P_i(Z_1 = j_1, \dots, Z_n = j_n) \frac{j_n q^{j_n}}{\gamma^n i q^i},$$

we arrive in the limit to a Markov chain with the transition probabilities $Q_n(i,j)$.

By Theorem 3 from Chap. I.11 in [1],

$$\gamma^{-n} P_n(i,j) \to i q^{i-1} \nu_j, \quad i,j \geq 1,$$

where $Q(s) = \sum_{j \geq 1} \nu_j s^j$ satisfies

$$Q(f(s)) = \gamma Q(s), \quad Q(q) = 0.$$

In the critical case as well as in the subcritical case with $\sum_{k=2}^{\infty} p_k k \log k = \infty$ the solution is trivial: $Q(s) \equiv 0$. Otherwise, $Q(s)$ is uniquely defined by the above equation with an extra condition $Q'(q) = 1$, so that the Q-process has a stationary distribution given by

$$Q_n(i,j) \to j q^{j-1} \nu_j,$$

with

$$\sum_{j \geq 1} j q^{j-1} \nu_j s^j = s Q'(sq).$$

These facts concerning $Q(s)$ remain valid even in the non-regular case. It is easy see from (14.2) that for our family with $\theta \neq 0$ and $A > q$, the generating function

$$Q(s) = (A-s)^{-\theta} - (A-q)^{-\theta},$$

is determined by parameters (θ, A) and is independent of $a = \gamma$. Similarly, for $\theta = 0$ we have

$$Q(s) = \log \frac{A-s}{A-q}.$$

This leaves us with two cases when $A = q = 1$. In the critical Case 2 the answer is trivial: $Q(s) \equiv 0$. In the subcritical Case 1, we have $\gamma = a^{-1/\theta}$ and

$$(1-f(s))^{-\theta} + d = \gamma^{-\theta}[(1-s)^{-\theta} + d],$$

which yields

$$Q(s) = [(1-s)^{-\theta} + d]^{-1/\theta}.$$

From these calculations it follows, in particular, that for our family of branching processes, in all subcritical cases, the classical $x \log x$ moment condition holds:

$$\sum_{k=2}^{\infty} p_k k \log k < \infty.$$

Using these explicit formulas for $Q(s)$ we can easily find the conditional probability distributions

$$\lim_{n\to\infty} P(Z_n = j | T > n) = b_j, \quad j \geq 1.$$

For all cases, except the critical Case 2, we have

$$\sum_{j\geq 1} b_j s^j = 1 - \frac{Q(sq)}{Q(0)}.$$

Turning to the Case 2, recall that for any critical Galton-Watson process, there exists a limit probability distribution

$$\lim_{n\to\infty} P(Z_n = j | T_0 = n+1) = w_j, \quad j \geq 1,$$

such that

$$\sum_{j\geq 1} w_j s^j = \lim_{n\to\infty} \frac{f_n(sp_0) - f_n(0)}{f_n(p_0) - f_n(0)}.$$

Since

$$f_n(sp_0) = 1 - [(1 - s(1 - [1 + c]^{-1/\theta}))^{-\theta} + nc]^{-1/\theta},$$

we obtain

$$\sum_{j\geq 1} w_j s^j = \frac{[1 - s(1 - [1 + c]^{-1/\theta})]^{-\theta} - 1}{c}.$$

14.7 Embedding into Continuous Time Branching Processes

Recall that a Galton-Watson processes with generating functions f_n is called *embeddable*, if there is a semigroup of probability generating functions

$$F_{t+u}(s) = F_t(F_u(s)), \quad t \in [0,\infty), u \in [0,\infty), \tag{14.3}$$

such that $f_n(s) = F_n(s)$, $n = 1, 2, \ldots$. Although not every Galton-Watson process is embeddable, see Chap. III.6 in [1], in this section we demonstrate that all theta-branching processes are embeddable.

Behind each semigroup (14.3) there is a continuous time Markov branching process with particles having exponential life lengths with parameter, say, λ. Each particle at the moment of death is replaced by a random number of new particles having a probability generating function

$$h(s) = h_0 + h_2 s^2 + h_3 s^3 + \ldots.$$

For such a continuous time branching process $(Z_t)_{t\in[0,\infty)}$ the probability generating function $F_t(s) = Es^{Z_t}$ satisfies

$$\int_s^{F_t(s)} \frac{dx}{h(x) - x} = \lambda t \tag{14.4}$$

(see [6] for a recent account of continuous time Markov branching processes). Our task for this section is for each $f \in \mathcal{G}_\theta$ to find a pair (h, λ) such that $f(s) = F_1(s)$. We will denote by $\mu = \sum_{k=2}^{\infty} kh_k$ the corresponding offspring mean number and by q the minimal nonnegative root of the equation $h(s) = s$ which gives the extinction probability of the continuous time branching process.

Cases 1–3 For a pair $\theta \in (0, 1]$ and $\mu \in (0, 1 + \theta^{-1}]$, put

$$h(s) = 1 - \mu(1 - s) + \frac{\mu}{1 + \theta}(1 - s)^{1+\theta}.$$

Taking successive derivatives of h it easy to see that it is a probability generating function with $h'(0) = 0$. Next we show that using this h as the offspring probability generating function for the continuous time branching process we can recover $f(s)$ for the theta-branching processes as $F_1(s)$ by choosing μ and λ adapted to Cases 1–3.

Case 1. For a given pair $a \in (0, 1)$ and $d \in (0, \infty)$, put

$$\mu = \frac{(1 + \theta)d}{(1 + \theta)d + 1}, \qquad \lambda = [(1 + \theta^{-1})d + \theta^{-1}] \ln a.$$

In this subcritical case, applying (14.4) we obtain for $s \in [0, 1)$

$$\lambda t = \int_s^{F_t(s)} \frac{dx}{(1 - \mu)(1 - x) + \frac{\mu}{1+\theta}(1 - x)^{1+\theta}} = \int_s^{F_t(s)} \frac{d \log \frac{1}{1-x}}{1 - \mu + \frac{\mu}{1+\theta} e^{\theta \log(1-x)}},$$

yielding the desired formula

$$F_t(s) = 1 - \Big(a^t(1 - s)^{-\theta} + (a^t - 1)d\Big)^{-1/\theta}.$$

Case 2. For a given $c \in (0, \infty)$, put $\mu = 1$ and $\lambda = (1 + \theta^{-1})c$. Then by (14.4), we get

$$F_t(s) = 1 - \Big((1 - s)^{-\theta} + ct\Big)^{-1/\theta}.$$

Case 3. If $\mu > 1$, then $q = 1-(\frac{(\mu-1)(1+\theta)}{\mu})^{1/\theta}$ and the proposed h can be rewritten as

$$h(s) = s + \frac{(1 - s)^{1+\theta} - (1 - q)^{\theta}(1 - s)}{1 + \theta - (1 - q)^{\theta}}.$$

For a given pair $a \in (0, 1)$ and $q \in [0, 1)$ choosing

$$\lambda = [(1 + \theta^{-1})(1 - q)^{-\theta} - \theta^{-1}] \ln a^{-1}$$

and applying (14.4), we obtain

$$F_t(s) = 1 - [a^t(1 - s)^{-\theta} + (1 - a^t)(1 - q)^{-\theta}]^{-1/\theta}.$$

It is easy to see that $f(s) = F_1(s)$ covers the whole subfamily $\mathcal{G}_\theta$ corresponding to the Cases 1–3.

Notice that if $\theta = 1$, then $h(s) = 1 - \frac{\mu}{2} + \frac{\mu}{2}s^2$ generates the linear birth and death process with $h''(1) = \mu$. If $\theta \in (0, 1)$, then $h''(1) = \infty$.

Case 4 Consider a supercritical reproduction law with infinite mean

$$h(s) = s + (1-s)\frac{\ln(1-s) - \ln(1-q)}{1 - \ln(1-q)}.$$

For $h_0 \in [0, 1)$ this can be rewritten as

$$h(s) = h_0 + (1-h_0)\sum_{k=2}^{\infty}\frac{s^k}{k(k-1)}.$$

In this form with $h_0 = 0$, the generating function h appeared in [4] as the reproduction law of an immortal branching process. Earlier in [8], this reproduction law was introduced as

$$h(s) = 1 - (1-h_0)(1-s)(1 - \ln(1-s)).$$

To see that the theta-branching process in the Case 4 is embeddable into the Markov branching process with the above mentioned reproduction law, use the first representation of h and apply (14.4). As a result we obtain for $s \neq q$,

$$\frac{\lambda t}{1 - \ln(1-q)} = \int_s^{F_t(s)} \frac{dx}{(1-x)(\ln(1-x) - \ln(1-q))} = \int_s^{F_t(s)} \frac{\ln(1-x)}{\ln(1-q) - \ln(1-x)}$$
$$= \ln[\ln(1-s) - \ln(1-q)] - \ln[\ln(1-F_t(s)) - \ln(1-q)].$$

Putting $\lambda = (1 - \ln(1-q))\ln a^{-1}$, we derive

$$F_t(s) = 1 - (1-q)^{1-a^t}(1-s)^{a^t}.$$

Cases 5, 7, 9. In these three cases the corresponding h and λ are given by an extension of the formulas for the Case 3:

$$h(s) = s + \frac{(A-s)^{1+\theta} - (A-q)^{\theta}(A-s)}{(1+\theta)A^{\theta} - (A-q)^{\theta}}, \quad \lambda = [(1+\theta^{-1})A^{\theta}(A-q)^{-\theta} - \theta^{-1}]\ln a^{-1}.$$

Turning to Definition 14.4 we see that this h in the Case 7 is dual to the h in the Case 3, and in the Case 9 it is dual to that of the Case 5.

Case 6. In this trivial case the corresponding continuous time branching process is a simple death-explosion process with $h(s) = q$ and $\lambda = \ln a^{-1}$.

Case 8. Similarly to the Case 4 we find that the pair

$$h(s) = s + (A-s)\frac{\ln(A-s) - \ln(A-q)}{1 + \ln A - \ln(A-q)}, \quad \lambda = (1 + \ln A - \ln(A-q))\ln a^{-1},$$

lead to

$$F_t(s) = A - (A-q)^{1-a^t}(A-s)^{a^t}.$$

Observe that this h is dual to that of the Case 4.

Acknowledgements This research was supported by the Swedish Research Council grant 621-2010-5623.

References

1. Athreya, K.B., Ney, P.E.: Branching Processes. Springer, Berlin (1972)
2. Harris, T.E.: The Theory of Branching Processes. Springer, Berlin (1963)
3. Klebaner, F., Rösler, U., Sagitov, S.: Transformations of Galton-Watson processes and linear fractional reproduction. Adv. Appl. Probab. **39**, 1036–1053 (2007)
4. Lagerås, A.N., Martin-Löf, A.: Genealogy for supercritical branching processes. J. Appl. Probab. **43**(4), 1066–1076 (2006)
5. Sagitov, S.: Linear-fractional branching processes with countably many types. Stoch. Process. Appl. **123**, 2940–2956 (2013)
6. Sagitov, S.: Tail generating functions for extendable branching processes. Preprint arXiv: 1511.05407 (2015)
7. Sagitov, S., Serra, M.C.: Multitype Bienayme-Galton-Watson processes escaping extinction. Adv. Appl. Probab. **41**, 225–246 (2009)
8. Sevastianov, B.A.: Branching Processes. Nauka, Moscow (1971) (in Russian)
9. Tokarev, D.: Galton-Watson processes and extinction in population systems. Ph.D. thesis, Monash University (2007)
10. Zolotarev, V.M.: More exact statements of several theorems in the theory of branching processes. Theory Prob. Appl. **2**, 245–253 (1957)

Part VII
Applications in Epidemiology

Chapter 15
Total Progeny of Crump-Mode-Jagers Branching Processes: An Application to Vaccination in Epidemic Modelling

Frank Ball, Miguel González, Rodrigo Martínez,
and Maroussia Slavtchova-Bojkova

Mathematics Subject Classification (2000): 60J80, 92D30

15.1 Introduction

In an epidemic context, the total number of infected individuals in a population is a useful tool for public health authorities in order to determine the infection level of a disease. Clearly, the total number of individuals that are infected is a key measure of the impact of an epidemic on the population within which it is spreading. Further, from an inferential viewpoint, surveillance systems typically provide more reliable information on the total number of individuals infected than on the precise temporal pattern of spread of an epidemic, so statistical analysis is often based on total infec-

---

F. Ball (✉)
School of Mathematical Science, The University of Nottingham, Nottingham NG7 2RD, UK
e-mail: frank.ball@nottingham.ac.uk

M. González
Department of Mathematics, University of Extremadura, Avda. Elvas, 06006 Badajoz, Spain
e-mail: mvelasco@unex.es

R. Martínez
Centro Universitario de Plasencia, Department of Mathematics, University of Extremadura, Plasencia, Spain
e-mail: rmartinez@unex.es

M. Slavtchova-Bojkova
Department of Probability, Operations Research and Statistics, Faculty of Mathematics and Informatics, Sofia University, Sofia, Bulgaria

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
e-mail: bojkova@fmi.uni-sofia.bg

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_15

tion data. This chapter is concerned with the study of the total size of an outbreak for epidemic models of diseases which follow an SIR (Susceptible–Infectious–Recovered) scheme in a closed, homogenously mixing population or some of its extensions, for example an SEIR (Susceptible-Exposed-Infectious-Recovered) scheme. When the population is homogeneously mixing and the number of infected individuals is small in relation to the total size of the susceptible population, it is well known that the number of infected individuals in such an epidemic may be well approximated by a single-type branching process, at least during its early stages (see, for example, [6, Chap. 3]). This approximation has a long history going back to the pioneering works of Bartlett [4] and Kendall [11], and can be made mathematically precise by considering a sequence of epidemics, indexed by the population size, and showing convergence of the process of infected individuals to a branching process as the population size tends to infinity (see, for example, [1], where such convergence of a very general epidemic model to a Crump–Mode–Jagers (CMJ) branching process—see [9]—is proved). Hence, we model the epidemic as a CMJ branching process, $Z = \{Z(t) : t \geq 0\}$, where $Z(t)$ denotes the number of infected individuals at time t. Thus $Z(0)$, which we assume to be fixed, represents the number of infected individuals at the beginning of the outbreak. Throughout the chapter, we assume that Z is non-explosive, i.e., that $\mathrm{P}(Z(t) < \infty) = 1$ for any $t \in (0, \infty)$. Conditions which guarantee this property may be found in [9, Sect. 6.2].

A key tool in controlling the spread of an epidemic is vaccination and there have been numerous mathematical studies of the effect of vaccination on disease dynamics. The majority of such studies using stochastic models have been concerned with the situation where a specified fraction of the population is vaccinated prior to an outbreak, though see [10, Chap. 8], for examples of analysis of more general vaccination policies in a deterministic setting. Recently, Ball et al. [3] have developed a framework for analysing time-dependent vaccination policies for epidemics which are modelled by a CMJ branching process. More specifically, a vaccination process is described by a function $\alpha : [0, \infty) \to [0, 1]$, such that $\alpha(t)$ represents the proportion of the population which is immune at time t ($t \geq 0$). Thus, since the population is homogeneously mixing, the probability that a contact at time t is with a non-immune individual is $1 - \alpha(t)$. (Modelling an epidemic as a CMJ branching process implies implicitly that changes in the susceptible population owing to infection of individuals are ignored.) For perfect vaccines, i.e. ones which confer lifelong immunity immediately with probability one, $\alpha(t)$ is given by the proportion of the population that has been vaccinated (i.e. the vaccination coverage) by time t. For imperfect vaccines, the vaccination coverage is implicitly included in the function α. For example, if the vaccine is all-or-nothing (i.e., it renders the vaccinee completely immune with probability ε_I, otherwise it has no effect), then $\alpha(t) = \varepsilon_I \tilde{\alpha}(t)$, where $\tilde{\alpha}(t)$ is the vaccination coverage at time t. Note that α is necessarily nondecreasing in t if the immunity conferred by vaccination does not wane.

Given a CMJ branching process Z and a vaccination process α, we denote by $Z_\alpha = \{Z_\alpha(t) : t \geq 0\}$ the vaccinated version of Z, in which each birth in Z is aborted independently, with probability $\alpha(t)$ if the birth time is at time t. Note that if a birth in Z is aborted in Z_α, then none of the descendants in Z of the aborted individual appear in Z_α. Hence, coupled realizations of Z and Z_α may be constructed by *pruning*, i.e. deleting individuals in Z and all of their descendants. In [3], such coupling was used to prove stochastic monotonicity and continuity properties, with respect to the vaccination process α, for functions defined on a CMJ branching process, first in a general context, i.e. for generic functions, and then specialized to the extinction time. However, these properties have not yet been explicitly obtained for the total progeny. Thus, in this chapter we establish explicitly these properties for the total number of infected individuals of the epidemic. To this end, we apply the general results given in [3], since total progeny is monotonically decreasing with pruning. In Sect. 15.2, we deduce the monotonicity and continuity properties of the mean and quantiles of the total progeny. After that, in Sect. 15.3, we present a simulated example, showing how to obtain in practice optimal vaccination policies (based on the results given in the previous section) to control the spread of a disease. The example is motivated by an outbreak of avian influenza virus in humans that occurred in Indonesia in 2006.

We end the introduction by describing some notation that will be used in the sequel. Let $\mathcal{A}$ be the space of all functions $\alpha : [0,\infty) \to [0,1]$. For any $c \in [0,1]$ and any $t_0 \geq 0$, we define the function $\alpha_c^{t_0} \in \mathcal{A}$ by

$$\alpha_c^{t_0}(t) = \begin{cases} 0 & \text{if } t < t_0, \\ c & \text{if } t \geq t_0, \end{cases}$$

which means that a proportion c of the population is vaccinated at time t_0. Thus, for example, α_c^0 denotes the constant function equal to c and α_0^0 denotes the constant function equal to 0. Moreover, let $c_{\text{inf}} = \max(0, 1 - m^{-1})$, where m is the offspring mean for Z. Thus, if $m > 1$, then c_{inf} is the critical vaccination coverage, i.e. the minimum proportion of the population that should be vaccinated with a perfect vaccine (at one single time) for the process to become critical, that is to assure the epidemic will go extinct with probability one. Note that the offspring mean of $Z_{\alpha^0_{c_{\text{inf}}}}$ is less than or equal to 1.

Also, for $t_0 \geq 0$ and $c \in [0,1]$, let

$$\mathcal{A}(c, t_0) = \{\alpha \in \mathcal{A} : \alpha(t) \geq c \text{ for all } t \geq t_0\}.$$

Finally, for $\alpha, \alpha' \in \mathcal{A}$, write $\alpha \prec \alpha'$ if $\alpha(t) \leq \alpha'(t)$ for all $t \in [0,\infty)$ and let $\|\alpha - \alpha'\| = \sup_{t\in[0,\infty)} |\alpha(t) - \alpha'(t)|$.

15.2 Monotonicity and Continuity Properties of Total Number of Infected Individuals Depending on Vaccination

For a given vaccination strategy, $\alpha \in \mathcal{A}$, we denote by $N_{\alpha,z}(\infty)$ the total number of births of the process Z_α when $Z(0) = z$, with $z \geq 1$, that is the total number of infected individuals in an outbreak when the vaccination process is defined by α.

This random variable is discrete, and in particular cases its probability distribution can be derived. For example, when the proportion of immune individuals in the population is constant with time, $N_{\alpha,z}(\infty)$ follows the Borel-Tanner distribution in some situations (see [8]). However, when a vaccination policy is time dependent, it is not easy to obtain the probability distribution of $N_{\alpha,z}(\infty)$ in a closed form.

Next we study monotonicity and continuity properties of the mean and quantiles of the total number of infected individuals depending on the vaccination function α. To this end, since individuals infect independently of each other, we have that

$$N_{\alpha,z}(\infty) = N^{(1)}_{\alpha,1}(\infty) + N^{(2)}_{\alpha,1}(\infty) + \ldots + N^{(z)}_{\alpha,1}(\infty),$$

where $N^{(i)}_{\alpha,1}(\infty)$ $(i = 1, 2, \ldots, z)$ are independent random variables with the same distribution as $N_{\alpha,1}(\infty)$. Hence

$$\mathrm{E}[N_{\alpha,z}(\infty)] = z\mu^N_\alpha,$$

where μ^N_α denotes the expectation of $N_{\alpha,1}(\infty)$, i.e. the mean number of infected individuals when the outbreak starts with one infected individual and the vaccination process is defined by α. Therefore, to analyze the behaviour of $\mathrm{E}[N_{\alpha,z}(\infty)]$, for any z, it is sufficient to study μ^N_α. Applying Theorem 3.1 and an obvious extension of Theorem 3.2(b) in [3], we deduce the following properties of μ^N_α.

Theorem 15.1

(a) If $\alpha, \alpha' \in \mathcal{A}$ satisfy $\alpha \prec \alpha'$, then $\mu^N_\alpha \geq \mu^N_{\alpha'}$.

(b) Fix $t_0 \geq 0$ and $c \in (c_{\text{inf}}, 1]$. Then, for each $\varepsilon > 0$, there exists $\eta = \eta(\varepsilon) > 0$ such that for all $\alpha, \alpha' \in \mathcal{A}(c, t_0)$ satisfying $\|\alpha - \alpha'\| \leq \eta$,

$$|\mu^N_\alpha - \mu^N_{\alpha'}| \leq \varepsilon. \tag{15.1}$$

Remark 15.1

(a) Notice that, under the conditions of Theorem 15.1, $\mu^N_{\alpha_c^{t_0}} < \infty$. Indeed, it is easy to obtain that, almost surely,

$$N_{\alpha_c^{t_0}}(\infty) \leq N(t_0) + \sum_{i=1}^{Z(t_0)} N^{(i)}_{\alpha_c^0,1}(\infty),$$

where $N(t_0)$ represents the total number of new infections in $(0, t_0)$. Therefore,

$$\mu^N_{\alpha_c^{t_0}} \leq \mathrm{E}[N(t_0)] + \mathrm{E}[Z(t_0)]\mu^N_{\alpha_c^0}.$$

The conditions which guarantee that the process is not explosive, imply that $\mathrm{E}[N(t_0)]$ and $\mathrm{E}[Z(t_0)]$ are finite. Moreover, $\mu^N_{\alpha_c^0}$ is also finite since $Z_{\alpha_c^0}$ is a subcritical process.

(b) If Z is subcritical, so $c_{\inf} = 0$, then Theorem 3.2(b) in [3] implies that for each $\varepsilon > 0$, there exists $\eta = \eta(\varepsilon) > 0$ such that (15.1) holds for all $\alpha, \alpha' \in \mathcal{A}$.

Notice that, in general, $\mathrm{P}(N_{\alpha,1}(\infty) \le \mu_\alpha^N) \ge 0.5$, because of the skewness of the distribution of the total progeny of the vaccinated CMJ branching process. Hence, if the vaccination policy α is applied, more than half of the outbreaks would have total size less than the mean μ_α^N, which may be sufficient protection for the population as a whole if the infectious disease is not too harmful for individuals. On the other hand, when the infectious disease is highly detrimental, we would like to control with high probability the total number of infected individuals and consequently consider vaccination policies based on quantiles of the total size distribution. Thus, fix $\alpha \in \mathcal{A}$, and define, for $0 < p < 1$,

$$x_{\alpha,p}^N = \inf\{x : \mathrm{P}(N_{\alpha,z}(\infty) \le x) \ge p\},$$

with the convention that $x_{\alpha,p}^N = \infty$ if $\mathrm{P}(N_{\alpha,z}(\infty) \le x) < p$ for all $x \in [0, \infty)$. Thus $x_{\alpha,p}^N$ is the quantile of order p of the random variable $N_{\alpha,z}(\infty)$. Since this variable is not derived from $N_{\alpha,1}(\infty)$, then the next result about the monotonicity and continuity properties of $x_{\alpha,p}^N$, deduced from Theorem 3.4 in [3], depends on z, for any $z > 0$.

Theorem 15.2 *Suppose that* $p \in (0, 1)$.

(a) If $\alpha, \alpha' \in \mathcal{A}$ *satisfy* $\alpha \prec \alpha'$, *then* $x_{\alpha,p}^N \ge x_{\alpha',p}^N$.
(b) Fix $t_0 \ge 0$ *and* $\alpha \in \mathcal{A}(c_{\inf}, t_0)$, *and let* $\{\alpha_n\}$ *be any sequence in* $\mathcal{A}$ *satisfying* $\alpha \prec \alpha_n$ *for all n and* $\lim_{n\to\infty} \|\alpha_n - \alpha\| = 0$. *Then* $\lim_{n\to\infty} x_{\alpha_n,p}^N = x_{\alpha,p}^N$.

Remark 15.2

(a) Notice that $Z_{\alpha_{c_{\inf}}^{t_0}}$ has offspring mean m until time t_0, and offspring mean $m_{c_{\inf}} \le 1$ after time t_0. Thus, since Z is non-explosive, the total progeny of $Z_{\alpha_{c_{\inf}}^{t_0}}$ is finite almost surely, and therefore $\mathrm{P}(N_{\alpha,z}(\infty) < \infty) = 1$ and $x_{\alpha,p}^N < \infty$, for all $\alpha \in \mathcal{A}(c_{\inf}, t_0)$ and $p \in (0, 1)$.
(b) Since $N_{\alpha,z}(\infty)$ is a discrete random variable for each α and z, then $x_{\alpha,p}^N$ is a step function depending on α.

Finally, from the above monotonicity and continuity properties of the mean and quantiles of $N_{\alpha,z}(\infty)$, and in the same way as it was described in Sect. 3.5 in [3], we propose how to choose optimal αs based on the total number of infected individuals. In particular, for fixed M, t_v and p_0, with $M \ge 0$, $0 \le p_0 \le 1$ and $0 \le t_v \le p_0^{-1}$, we define the function α_{M,t_v,p_0}, where, for $s \ge 0$,

$$\alpha_{M,t_v,p_0}(s) = \begin{cases} 0, & \text{if } s \le M \\ p_0(s - M), & \text{if } M < s \le M + t_v \\ t_v p_0, & \text{if } M + t_v < s. \end{cases} \qquad (15.2)$$

This function describes the proportion of immune individuals in the population when the vaccination process starts at time M, takes t_v time units and the proportion of individuals vaccinated per unit time is p_0. In practice, M and p_0 are usually known before vaccination begins, in which case the vaccination function α_{M,t_v,p_0} can be parameterized through t_v alone. Hence, for fixed M and p_0, let $\alpha_{t_v} = \alpha_{M,t_v,p_0}$ and $\mathcal{A}^* = \{\alpha_{t_v} : c_{\inf}p_0^{-1} \leq t_v \leq p_0^{-1}\}$. Since $\mathcal{A}^*$ is a subset of $\mathcal{A}(c_{\inf}, M + c_{\inf}p_0^{-1})$, then Theorems 15.1 and 15.2 ensure that, for each $b \geq 0$ and $p \in (0,1)$, optimal vaccination policies based on the mean and quantiles exist and are unique, provided that $\{\alpha \in \mathcal{A}^* : z\mu^N_{\alpha_{t_v}} \leq b\}$ and $\{\alpha \in \mathcal{A}^* : x^N_{\alpha_{t_v},p} \leq b\}$ are non-empty. Then, we denote by $\alpha_{t^N_{\text{opt},\mu,b}}$ and $\alpha_{t^N_{\text{opt},p,b}}$, the corresponding optimal policies based on the mean and quantiles, respectively, where

$$t^N_{\text{opt},\mu,b} = \inf\{t_v : z\mu^N_{\alpha_{t_v}} \leq b\} \quad \text{and} \quad t^N_{\text{opt},p,b} = \inf\{t_v : x^N_{\alpha_{t_v},p} \leq b\}.$$

Notice that these optimal policies depend on M and p_0, which have been fixed previously. Moreover, $x^N_{\alpha_{t^N_{\text{opt},p,b}},p} \leq b$, though equality is not guaranteed since $N_{\alpha_{t^N_{\text{opt},p,b}},z}(\infty)$ is a discrete random variable. On the other hand, $z\mu^N_{t^N_{\text{opt},\mu,b}} = b$, if $t^N_{\text{opt},\mu,b} > c_{\inf}p_0^{-1}$.

15.3 Simulated Example

To illustrate how to obtain optimal vaccination strategies based on the mean and quantiles of the total size of an outbreak, we present a simulation study which has been motivated by an outbreak of avian influenza in humans that occurred in Indonesia in 2006. The spread of this disease can be considered as an SEIR epidemic and therefore its early spread can be approximated by a CMJ branching process. In our simulations, we consider an offspring mean of 1.14, so the corresponding CMJ process is supercritical and hence a vaccination strategy (or some other mitigation measure) should be applied in order to control the outbreak. It is known (see [15]) that for the transmission of avian influenza in humans, the latent period (the period elapsing between infection of an individual and the beginning of his/her infectious period) has a probable range of 3–7 days and the infectious period has a probable range of 5–13 days. So, in our study we assume that the latent and infectious periods are independent random variables which follow gamma distributions with means 5 and 9, and shape parameters 23 and 19, respectively. Hence, approximately 95 % of incubation periods are between 3 and 7 days and approximately 95 % of infectious periods are between 5 and 13 days. Furthermore, we assume that during the infectious period, infections occur according to a homogeneous Poisson process, independently of the duration of incubation and infectious periods. Since the mean of the infectious period is 9 days and the offspring mean is 1.14, we assume that this Poisson process has rate $1.14/9 = 0.126667$. These kind of distributions are appropriate for latent and infectious periods, and for the number of infections made by a typical infective (see for example [5, 7, 8] or [12]). In [15], the spread of avian influenza in humans is modelled using a households epidemic model (see [2]) and the local (i.e. within-household) basic reproduction number R_0 is estimated to be 1.14. For homogeneously mixing epidemic models, R_0 is given by the offspring mean of the corresponding approximating CMJ branching process (see [1]). The definition of R_0 is more complicated for epidemic models with household structure (see [13]). Moreover, it is possible for the epidemic in the population at large to be subcritical when the local reproduction number is greater than one, and vice versa. Thus, although our choice of 1.14 for the offspring mean of the CMJ process used to model the early spread of the disease corresponds to the estimate of the local reproduction number in [15], it may well not reflect the true R_0 for a homogeneously mixing model of avian influenza in humans. Finally, all the simulations start with a single index patient.

Assuming the previous modelling and taking into account the kind of vaccination policies defined by (15.2), we seek an optimal vaccination strategy belonging to the set

$$\mathcal{A}^* = \{\alpha_{M,t_v,0.01} \in \mathcal{A} : M \in \mathbb{N} \cup \{0\},\ 13 \le t_v \le 100\},$$

where, for $s \ge 0$,

$$\alpha_{M,t_v,0.01}(s) = \begin{cases} 0, & \text{if } s \le M \\ 0.01(s-M), & \text{if } M < s \le M + t_v \\ 0.01 t_v, & \text{if } M + t_v < s. \end{cases}$$

We recall that M represents the number of days until the vaccination process starts, 0.01 (1 %) determines the proportion of individuals vaccinated per day during the vaccination process and t_v indicates the duration of the vaccination process (in days). Notice that 13 is the smallest value of t_v such that the vaccinated process becomes subcritical. On the other hand, when $t_v = 100$, all individuals are vaccinated during the vaccination process, minimizing the propagation of the virus.

The left plot in Fig. 15.1 shows the behaviour of $\mu^N_{\alpha_{M,100,0.01}}$, for $M \in \{0, 1, \ldots, 21\}$, which has been estimated by using the Monte-Carlo simulation-method described in Sect. 3.5 in [3]. Specifically, we have simulated 10,000 processes and from them we have estimated $\mu^N_{\alpha_{M,t_v,0.01}}$, for each $M \in \{0, 1, \ldots, 21\}$ and $t_v \in \{13 + k : k = 0, 1, \ldots, 87\}$. Notice that, by Theorem 15.1(a), when M increases the mean of the total number of infected individuals also increases. We observe that for M greater than 11 (vertical dotted line in the plot), the mean of the total number of new infected individuals in an outbreak started with one infected individual is greater than 7 (horizontal dotted line in the plot), the size of the outbreak detected in Indonesia (see [15]). Hence, for each $M \in \{0, 1, \ldots, 11\}$, the optimal duration of the vaccination

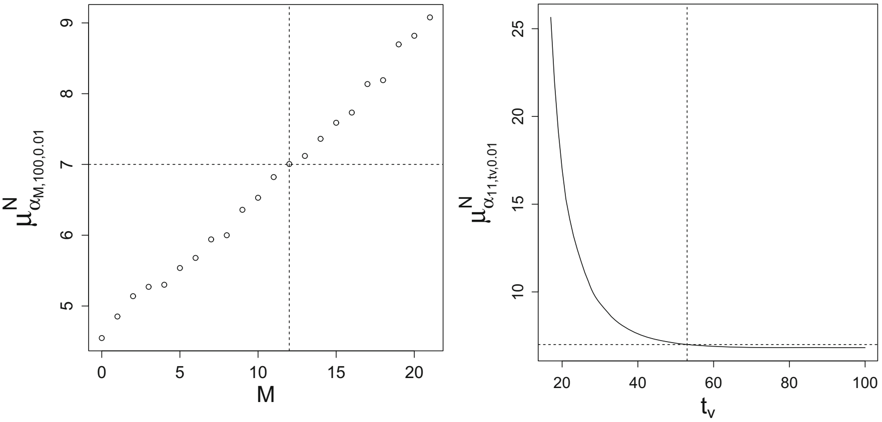


Fig. 15.1 *Left*: Behaviour of estimated value of $\mu^N_{\alpha_{M,100,0.01}}$ depending on M. *Right*: Behaviour of estimated value of $\mu^N_{\alpha_{11,t_v,0.01}}$ depending on t_v

Table 15.1 Estimated optimal duration of vaccination depending on M

M	0	1	2	3	4	5	6	7	8	9	10	11
$t^N_{opt,M,\mu,7}$	29	31	32	33	33	35	36	38	38	42	45	53

based on the mean, $t^N_{\text{opt},M,\mu,7}$, is given by

$$t^N_{\text{opt},M,\mu,7} = \inf\{t_v : \mu^N_{\alpha_{M,t_v,0.01}} \leq 7\}.$$

Table 15.1 shows the estimated optimal duration of vaccination depending on the number of days until the vaccination process starts. One can observe that these optimal durations increase when M increases.

Now, we focus our attention on M equal to 11 (the most unfavourable situation). The right plot in Fig. 15.1 shows the behavior of the estimates of $\mu^N_{11,t_v,0.01}$, which decrease in a continuous way when t_v increases, by Theorem 15.1. We find that the optimal duration of vaccination, again to guarantee a mean number of infected individuals no greater than 7, is 53 days (which means that at the end of the vaccination process, we have vaccinated 53 % of susceptible individuals), so vaccination ceases 64 days (more than 2 months) after the start of an outbreak. The left plot in Fig. 15.2 shows the distribution of the total number of infected individuals in the outbreak after applying this optimal vaccination procedure. Obviously, although the mean of this distribution is 7 (dotted line in the plot), there exists a positive probability that more than seven individuals are infected by the virus, which is less than 0.5, because of the skewness of the distribution. Indeed, $x^N_{\alpha_{11,53,0.01},0.703}$ is estimated by 7, that is, vaccinating 53 % of susceptible individuals guarantees that at least 70.3 % of new outbreaks infect no more than seven individuals. Table 15.2 quantifies this probability as well as the probabilities of being no greater than other

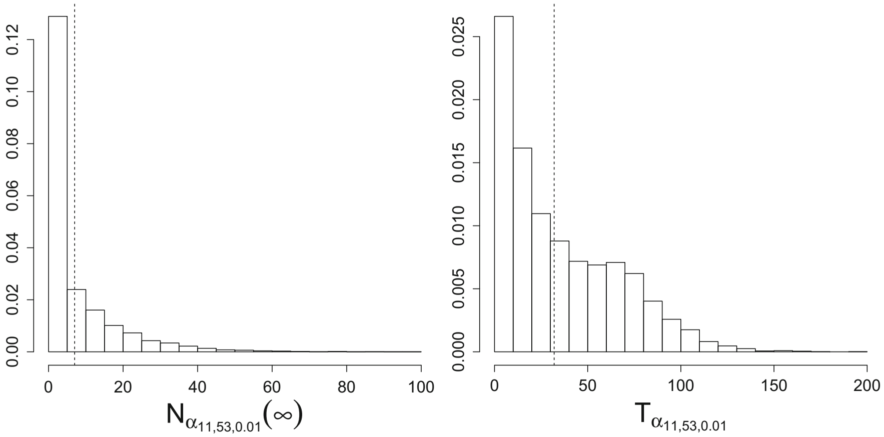


Fig. 15.2 *Left*: Histogram of simulated total number of new infected individuals for $M = 11$ and $t_v = 53$. *Right*: Histogram of simulated time to extinction for $M = 11$ and $t_v = 53$

Table 15.2 Estimated probabilities of the total number of infected individuals, after applying optimal vaccination policy $\alpha_{11,53,0.01}$

x	7	15	25	50	75	100
$P(N_{\alpha_{11,53,0.01}}(\infty) \leq x)$	0.703	0.845	0.932	0.992	0.998	1

upper bounds. The greater that probability is, the greater will be the total number of infected individuals. Hence, when the vaccination policy $\alpha_{11,53,0.01}$ is applied, more than 90 % of new outbreaks infect no more than 25 individuals. Moreover, the right plot in Fig. 15.2 shows the distribution of the time to extinction of outbreaks started with one infected individual when one applies this policy. We estimate that 55.7 % of the outbreaks become extinct before 32 days (dotted line in the plot), the observed value for the outbreak detected in Indonesia (see [15]).

For fixed $p = 0.90$, a high probability, $x^N_{\alpha_{11,t_v,0.01},0.90}$ decreases as a step function in t_v, according to Theorem 15.2. The left plot in Fig. 15.3 shows this behavior. Finally, notice that, although ultimately all susceptible individuals are vaccinated, at least 10 % of new outbreaks infect more than 20 individuals, since this particular vaccination strategy takes a time (100 days) to be completely applied (in general, because $M > 0$ and $p_0 \neq 1$). Moreover, since $t^N_{opt,0.90,20}$ is estimated by 64, we deduce that the optimal vaccination coverage is 64 %. The same behaviour is found for $p = 0.95$, where $t^N_{\text{opt},0.95,29}$ is estimated by 55, see right plot in Fig. 15.3.

Remark 15.3 For the computer simulations, we used the language and environment for statistical computing and graphics **R** ("GNU S") (see [14]).

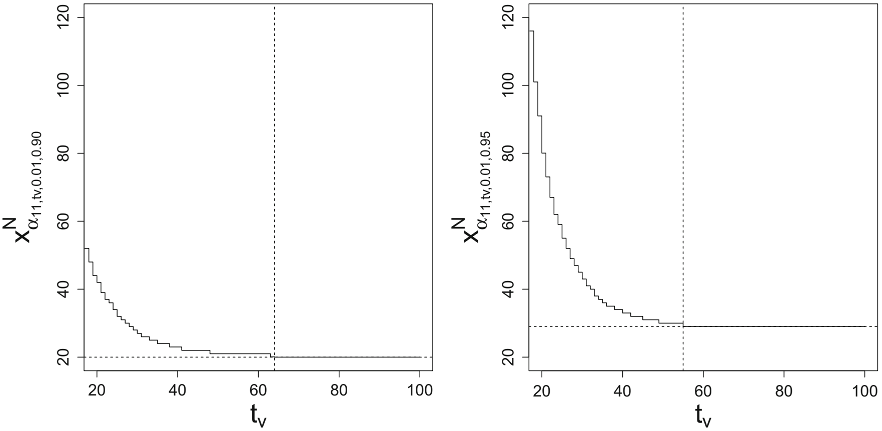


Fig. 15.3 *Left*: Behaviour of estimated value of $x^N_{\alpha_{11,t_v},0.01,0.90}$ depending on t_v. *Right*: Behaviour of estimated value of $x^N_{\alpha_{11,t_v},0.01,0.95}$ depending on t_v

Acknowledgements We thank the referee for carefully reading our chapter and for constructive comments which have improved its presentation. The research was supported by the Ministerio de Economía y Competitividad of Spain (grants MTM2012-31235 and MTM2015-70522-P), the Junta de Extremadura (grant GR15105), the FEDER, and also by National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, (grant NFSI NI02/17).

References

1. Ball, F., Donnelly, P.: Strong approximations for epidemic models. Stoch. Process. Appl. **55**(1), 1–21 (1995)
2. Ball, F., Mollison, D., Scalia-Tomba, G.: Epidemics with two levels of mixing. Ann. Appl. Probab. **7**(1), 46–89 (1997)
3. Ball, F., González, M., Martínez, R., Slavtchova-Bojkova, M.: Stochastic monotonicity and continuity properties of functions defined on Crump-Mode-Jagers branching processes, with application to vaccination in epidemic modelling. Bernoulli **20**(4), 2076–2101 (2014)
4. Bartlett, M.S.: An Introduction to Stochastic Processes. Cambridge University Press, Cambridge (1955)
5. Daley, D.J., Gani, J.: Epidemic Modelling: An Introduction. Cambridge University Press, Cambridge (1999)
6. Diekmann, O., Heesterbeek, H., Britton, T.: Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press, Princeton (2013)
7. Farrington, C., Grant, A.: The distribution of time to extinction in subcritical branching processes: applications to outbreaks of infectious disease. J. Appl. Probab. **36**(3), 771–779 (1999)
8. Farrington, C., Kanaan, M., Grant, A.: Branching process models for surveillance of infectious diseases controlled by mass vaccination. Biostatistics **4**(2), 279–295 (2003)
9. Jagers, P. Branching Processes with Biological Applications. Wiley, New York (1975)
10. Keeling, M.J., Rohani, P.: Modeling Infectious Diseases in Humans and Animals. Princeton University Press, Princeton (2008)

11. Kendall, D.G.: Deterministic and stochastic epidemics in closed populations. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 149–165. University of California Press, Berkeley (1956)
12. Mode, C.J., Sleemam, C.K.: Stochastic Processes in Epidemiology. World Scientific, Singapore (2000)
13. Pellis, L., Ball, F., Trapman P.: Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of R_0. Math. Biosci. **235**(1), 85–97 (2012)
14. R Development Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/ (2015). ISBN 3-900051-07-0
15. Yang, Y., Halloran, M., Sugimoto, J., Longini, I.: Detecting human-to-human transmission of avian influenza A (H5N1). Emerg. Infect. Dis. **13**(9), 1348–1353 (2007)

Chapter 16
Inference for Emerging Epidemics Among a Community of Households

Frank Ball and Laurence Shaw

Mathematics Subject Classification (2000): 60J80, 62M05, 92D30

16.1 Introduction

Branching processes play a key role in the analysis of epidemic models. In the early stages of an epidemic with few initial infectives in a large homogeneously-mixing population, the probability that an infectious contact is with a previously infected individual is very small, so the process of infectives is well approximated by a branching process. Such an approximation has a long history going back to Bartlett [8] and Kendall [15] and leads to the celebrated threshold theorem, which broadly states that an epidemic with few initial infectives can take off only if the initial susceptible population size is sufficiently large (see e.g. Whittle [18]). This theorem holds for very general homogeneously-mixing epidemic models (see Ball and Donnelly [5]) and has important practical applications, such as in determining the critical vaccination coverage to prevent the occurrence of a large epidemic.

The above results extend to multitype populations but they require the population to be locally, as well as globally, large, in the sense that if the population is partitioned into groups, for example, by age, sex and/or geographical location, then each of the groups, and not just the total population, must be large. However, such an assumption is clearly unrealistic for most epidemics in human populations, since such populations contain small groups, such as households, school classes

---

F. Ball (✉) • L. Shaw
School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK
e-mail: frank.ball@nottingham.ac.uk; pmxlmsha@nottingham.ac.uk.

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_16

and workplaces, in which transmission is likely to be enhanced. Consequently, there has been a growing interest in epidemics among structured populations. One approach is to develop complex simulation models; these have been widely used to inform public health policy but can be difficult to interpret owing to their complexity. Another approach is to develop models which capture key departures from homogeneous mixing but are still analytically tractable. One such class of models is *household* models (e.g. Becker and Dietz [9] and Ball et al. [7]), which assume that the population is partitioned into households, with different transmission rates for within- and between-household infections. Household structure is clearly highly relevant for human populations and has significant impact on disease dynamics and the performance of vaccination strategies. Further, epidemic data on emerging infections are often collected at a household level (e.g. Cauchemez et al. [10] and House [13]).

In order to evaluate the efficacy of mitigation strategies for an emerging epidemic it is necessary to have estimates of parameters governing an epidemic model. For an emerging homogeneously-mixing epidemic, Wallinga and Lipsitch [17] provided a simple estimate of the basic reproduction number R_0 (see e.g. Heesterbeek and Dietz [12]) based on knowledge of the early exponential growth rate r of the epidemic and the generation interval for the disease. (In branching process terminology, R_0 is the offspring mean, r is the Malthusian parameter and knowing the generation interval corresponds to knowing the reproduction kernel up to a multiplicative constant.) Fraser [11] extended this methodology to household models but his method requires knowledge of within-household transmission parameters, which may well be unknown for a new emerging infection. Estimation of such parameters is not straightforward as the emerging nature of the epidemic needs to be accounted for correctly.

In a recent paper, Ball and Shaw [6] used the asymptotic theory of CMJ (Crump-Mode-Jagers) branching processes, developed by Nerman [16], to provide an estimator of the within-household infection rate from data collected during an emerging infection; specifically, it was assumed that an estimate of the early exponential growth rate r is available from general data on an emerging epidemic and more-detailed household-level data are available in a sample of households. The method is easily implemented numerically only for Markov epidemics, where the infectious period follows an exponential distribution. For such epidemics, an alternative but equivalent approach is via the asymptotic theory of continuous-time, multitype Markov branching processes. The purpose of this chapter is to present that approach and also, for Markov epidemics, a new way of calculating both the total size distribution for a single-household epidemic and the threshold parameter R_* for the household model. For ease of exposition, we present the results within the framework of a population of equally-sized households. Extension to unequal household sizes is straightforward, cf. Ball and Shaw [6].

The remainder of the chapter is organised as follows. The household model, its approximation by a CMJ branching process and associated threshold parameter R_* are described in Sect. 16.2. Estimation during the early stages of an emerging epidemic is considered in Sect. 16.3. A basic method, which ignores the emerging nature of an epidemic is described in Sect. 16.3.1, where numerical illustrations show that it can lead to biased estimates. The method using a multitype Markov branching process is given in Sect. 16.3.2, together with simulations that show that the method is feasible for realistic population sizes. For ease of exposition, we assume that all infected households are observed; modification to the case when only a sample of households is observed is straightforward, provided that the sampling does not induce bias. A modified single-household epidemic process is described in Sect. 16.4, where it is used to determine R_* and the single-household epidemic total size distribution, and also to show that the basic method becomes approximately unbiased as $r \downarrow 0$. Some brief concluding comments are given in Sect. 16.5.

16.2 Household Model and Threshold Parameter

16.2.1 Model

Consider a closed population of N $(= mn)$ individuals, who reside in m households each having size n. The epidemic is initiated by a small number individuals becoming infected at time $t = 0$, with the other individuals in the population all assumed to be susceptible. The infectious periods of infectives are distributed according to a random variable I, having an arbitrary but specified distribution. We assume that $\mathrm{E}[I] = 1$, which, provided $\mathrm{E}[I] < \infty$, involves no loss of generality as time can be linearly rescaled accordingly. Throughout its infectious period, a given infective makes *global* contacts with any given susceptible in the population at the points of a homogeneous Poisson process having rate $\frac{\lambda_G}{N}$ and, additionally, it makes *local* contacts with any given susceptible in the same household at the points of a homogeneous Poisson process having rate λ_L. All the Poisson processes describing infectious contacts (whether or not either or both of the individuals involved are the same) and the random variables describing the infectious periods are mutually independent. A susceptible becomes an infective as soon as it is contacted by an infective (i.e. there is no latent period). An infective recovers at the end of its infectious period and plays no further role in the epidemic. The epidemic terminates when there is no infective remaining in the population.

16.2.2 Threshold Parameter R_*

Suppose that the number of households m is large and the initial number of infectives is small. Then, during the early stages of an epidemic, the probability that a global contact is with an individual in a previously infected household is very small. Thus the initial behaviour of the epidemic can be approximated by a CMJ branching process, describing the proliferation of infected households, in which all global contacts are assumed to be with individuals who reside in fully susceptible households. This approximation can be made exact by letting m tend to infinity, with n fixed, and using a coupling argument (Ball [4]). More precisely, a sequence of epidemic processes, indexed by m, and the limiting CMJ branching process can be constructed on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$ so that, if A_{ext} and r denote respectively the extinction set and Malthusian parameter of the branching process, then (i) for P-almost all $\omega \in A_{\text{ext}}$, the branching process and the process of infected households in the epidemic process coincide for all sufficiently large m; and (ii) for any $c \in (0, (2r)^{-1})$, for P-almost all $\omega \in \Omega \backslash A_{\text{ext}}$ these processes coincide throughout the time interval $[0, c \log m]$ for all sufficiently large m. Let R_* be the mean number of global contacts that emanate from a typical single-household epidemic. Then R_* is a threshold parameter for the epidemic process, in that as $m \to \infty$, an epidemic with few initial infectives takes off with non-zero probability if and only if $R_* > 1$ (Ball et al. [7]).

To determine R_*, consider a single-household epidemic, with initially one infective, labelled 0, and $n-1$ susceptibles, labelled $1, 2, \cdots, n-1$, and ignore the effect of global contacts. For $i = 0, 1, \cdots, n-1$, let I_i denote the length of individual i's infectious period, if i were to be infected, and let $\chi_i = 1$ if i is infected by the single-household epidemic and $\chi_i = 0$ otherwise. Then, as $m \to \infty$, the total number of global contacts that would emanate from this household, if allowed, follows a Poisson distribution with random mean $\lambda_G A$, where $A = \sum_{i=0}^{n-1} \chi_i I_i$, whence $R_* = \lambda_G \mathrm{E}[A]$. Further, by exchangeability and noting that, for each $i = 0, 1, \cdots, n-1$, χ_i and I_i are independent,

$$R_* = \lambda_G \mathrm{E}[Z]\mathrm{E}[I],$$

where $Z = \chi_0 + \chi_1 + \cdots, \chi_{n-1}$ is the total size (i.e. total number of individuals infected, including the initial infective) of the single-household epidemic.

Let $p_{\text{basic}}(i|\lambda_L) = \mathrm{P}(Z = i)$ $(i = 1, 2, \cdots, n)$. Then (Ball [3, Eq. (2.5)]),

$$\sum_{i=1}^{j} \binom{n-i}{j-i} \frac{p_{\text{basic}}(i|\lambda_L)}{\phi((n-j)\lambda_L)^i} = \binom{n-1}{j-1} \quad (j = 1, 2, \cdots, n),$$

where $\phi(\theta) = \mathrm{E}[\exp(-\theta I)]$ $(\theta \geq 0)$ is the moment-generating function of I. This triangular system of linear equations enables $p_{\text{basic}}(i|\lambda_L)$ $(i = 1, 2, \cdots, n)$, and hence also $\mathrm{E}[Z]$, to be calculated. An alternative method of calculating these quantities when I has an exponential distribution is presented in Sect. 16.4.

16.3 Estimation in the Early Stages of an Epidemic

16.3.1 Basic Method

Suppose one wishes to estimate λ_L for an epidemic that is observed whist it is still in its initial stages and is therefore still mimicking the infected households branching process outlined above. Assuming that individuals in each of the three possible states (susceptible, infective and recovered) are distinguishable from one another, let $a_{x,y}$ be the number of households in state (x, y), i.e. containing x susceptibles and y infectives, at the time when the epidemic is observed. By considering only the households in which the single-household epidemic has ceased (i.e. where $x < n$ and $y = 0$), one can attempt to estimate λ_L by maximising the pseudolikelihood function

$$L_{\text{basic}}(\lambda_L|\boldsymbol{a}) = \prod_{i=1}^{n} p_{\text{basic}}(i|\lambda_L)^{a_{n-i,0}}.$$

Note that L_{basic} is not a true likelihood as it assumes the epidemics in distinct households are independent. We call this method *basic MPLE*.

This method of estimation is simple but does not use all of the information available since households in which infectives are still present are not used. A similar approach using all of the information available is to use maximum pseudolikelihood estimation but with censoring on households in which there are still infectives remaining. For $i = 1, 2, \cdots, n$, let $q_{\text{basic}}(i|\lambda_L) = \sum_{j=i}^{n} p_{\text{basic}}(j|\lambda_L)$ be the probability that the total size of a single-house epidemic is at least i and let $b_i = \sum_{y=1}^{i} a_{n-i,y}$ be the number of observed households with $n - i$ susceptibles and at least one infective. The total size of an epidemic in such a household is at least i, so λ_L may be estimated by maximising the (right-censored) pseudolikelihood function:

$$L_{\text{censor}}(\lambda_L|\boldsymbol{a}, \boldsymbol{b}) = \prod_{i=1}^{n} \left[p_{\text{basic}}(i|\lambda_L)^{a_{n-i,0}} q_{\text{basic}}(i|\lambda_L)^{b_i} \right].$$

This method, which we call *censored MPLE*, assumes that if a household is in state (x, y), with $y \geq 1$, when the epidemic is observed then the total size of the epidemic in that household is at least $z + y$, where $z = n - x - y$ is the number of recoveries in the household when estimation is made. An alternative method, called *censored MPLE (recoveries only)*, assumes instead that the total size of the epidemic in such a household is at least $z + 1$.

The performance of the basic and censored MPLE methods is illustrated in Fig. 16.1, which considers simulated epidemics, with infectious period I following an exponential distribution with mean one and infection parameters $\lambda_G = 1.21$ and $\lambda_L = 0.64$, initiated by a single infective in a population of $m = 1{,}000{,}000$ households, each having size 4. For each simulated epidemic that took off, estimates

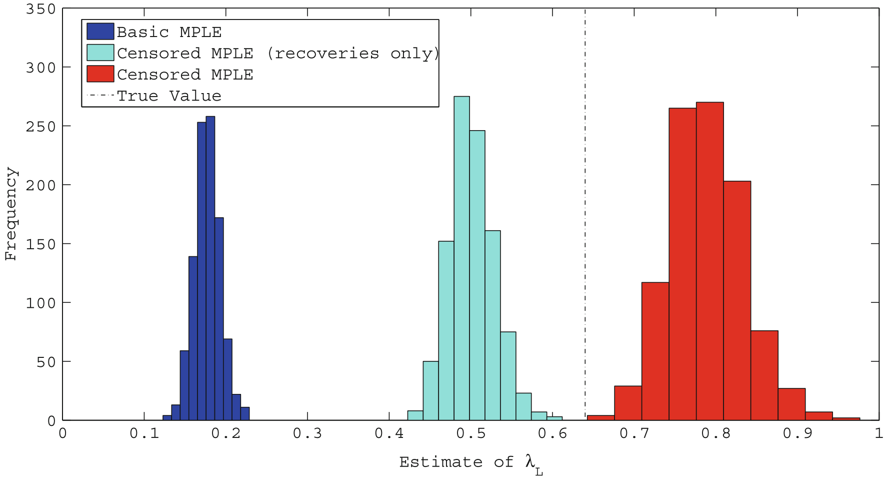


Fig. 16.1 Histograms of estimates of λ_L (true value 0.64 shown by *vertical line*) from 1000 simulated epidemics that took off using the basic and censored MLE methods outlined above

of λ_L were made after 1000 recoveries had occurred using the three methods described above. Any epidemic not reaching 1000 recoveries was deemed not to have taken off. A large value of m was used to ensure that the branching process approximation of the epidemic is still accurate when estimation is made. The chosen values of λ_L and λ_G broadly reflect the influenza example in Fraser [11].

It is clear from Fig. 16.1 that the basic MPLE method yields a gross underestimate of λ_L. This is because estimation is based on completed single-household epidemics and smaller single-household epidemics are more likely than larger ones to have finished at the time when estimation is made. The censored MPLE methods perform better but note that the recoveries-only censored method still leads to an appreciable underestimate of λ_L, while the other censored method leads to an overestimate of λ_L. In order to obtain a more accurate estimate of λ_L, we exploit the asymptotic behaviour of the CMJ branching process which approximates the proliferation of infected households, introduced in Sect. 16.2.2, to determine the approximate distribution of the state of a typical single-household epidemic when estimation is made.

16.3.2 Multitype Birth-Death Process Method

16.3.2.1 The Infected Households Branching Process as a Multitype Birth-Death Process

Suppose that the infectious period I follows an exponential distribution with mean 1. Recall that individuals in the above-mentioned CMJ branching process correspond to single-household epidemics. Let the type of an individual be given by the state (x, y) of its associated single-household epidemic, where x is the number of susceptibles and y is the number of infectives in the household and note that the type of an individual in the branching process changes whenever a susceptible becomes infected or an infective recovers in the corresponding single-household epidemic. Define active individual types to be those in which $y \geq 1$. The approximating branching process can be described by a multitype birth-death (B-D) process, S, on the active individual types. Let $\mathcal{T} = \{(x, y) : x \geq 0, y \geq 1, x + y \leq n\}$ denote the type space of S.

Consider an individual in S of type (x, y). Such an individual has an exponentially distributed lifetime with rate $y(1 + x\lambda_L)$, during which it gives birth to type-$(n-1, 1)$ individuals at rate $y\lambda_G$ as a result of infectives making global contacts with susceptibles in previously uninfected households. Upon death, a type-(x, y) individual produces a type-$(x-1, y+1)$ individual with probability $x\lambda_L/(x\lambda_L + 1)$, otherwise it produces a type-$(x, y-1)$ individual if $y \geq 2$ or no individual if $y = 1$, since the recovery of the last remaining infective in a household causes a single-household epidemic to cease. Note that the type space $\mathcal{T}$ has $k = n(n+1)/2$ elements. Label the types $1, 2, \cdots, k$, such that for $i = 1, 2, \cdots, n$, a type-$(n-i, 1)$ individual has label i. Figure 16.2 gives a graphical representation of a single-household epidemic as a multitype B-D process for $n = 3$, using such a labelling system for individual types.

Let $\Lambda = [\lambda_{ij}]$ be the $k \times k$ birth-rate matrix of S, with λ_{ij} being the rate at which a type-i individual gives birth to a type-j individual, and let $\text{diag}(\boldsymbol{\mu})$ be the $k \times k$ diagonal death-rate matrix of the process, with consecutive diagonal elements given by $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_k)$, where μ_i is the rate at which a type-i individual dies. (Thus if i corresponds to the state(x, y), $\mu_i = y(1 + x\lambda_L)$, $\lambda_{ij} = y\lambda_G$ if

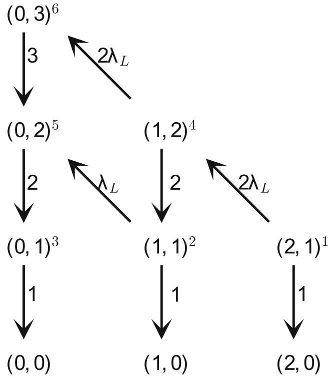


Fig. 16.2 Graphical representation of a single-household epidemic as a multitype B-D process for households of size 3, where (x, y) denotes the household state and type labels (shown as superfixes) for the multitype B-D process are assigned as described in Sect. 16.3.2.1. The values on the arrows represent transition rates between household states in the epidemic and the rate at which individuals give birth to individuals of other types in the B-D process. Note that all labelled types also give birth to individuals labelled as type-1 at rate $y\lambda_G$ in the full model

j corresponds to the state $(n-1,1)$, $\lambda_{ij} = xy\lambda_L$ if j corresponds to the state $(x-1, y+1)$, $\lambda_{ij} = y$ if j corresponds to the state $(x, y-1)$ and $\lambda_{ij} = 0$ for all other j.) Let $r(\lambda_L, \lambda_G)$ be the maximal eigenvalue of $A = \Lambda - \text{diag}(\boldsymbol{\mu})$. Then $r(\lambda_L, \lambda_G)$ is the Malthusian parameter of S and hence is also the early exponential growth rate of the households epidemic, which we assume is observable. Let $\boldsymbol{v}(\lambda_L, \lambda_G) = (v_1(\lambda_L, \lambda_G), v_2(\lambda_L, \lambda_G), \cdots, v_k(\lambda_L, \lambda_G))$ be the left-eigenvector of A associated with $r(\lambda_L, \lambda_G)$, normalised such that $\sum_{i=1}^{k} v_i(\lambda_L, \lambda_G) = 1$.

For $t > 0$, let $\boldsymbol{Y}(t) = (Y_1(t), Y_2(t), \cdots, Y_k(t))$ denote the number of individuals of each type from $\mathcal{T}$ alive at time t in S. It follows from Athreya and Ney [2], page 206, that

$$\mathrm{e}^{-r(\lambda_L,\lambda_G)t}\boldsymbol{Y}(t) \xrightarrow{a.s.} W\boldsymbol{v}(\lambda_G, \lambda_L) \quad \text{as } t \to \infty, \tag{16.1}$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence and $W \geq 0$ is a random variable such that $W = 0$ if and only if S becomes extinct. The normalised eigenvector $\boldsymbol{v}(\lambda_G, \lambda_L)$ therefore gives the proportions of individuals of each type in S as $t \to \infty$ when S does not go extinct.

16.3.2.2 Estimating $\boldsymbol{\lambda_L}$

The vector $\boldsymbol{v}(\lambda_L, \lambda_G)$ gives the proportions of each individual type in S still remaining but it does not give explicit information concerning single-household epidemics that have ceased before the epidemic is observed. For $i = 1, 2, \cdots, n$, let $\tilde{Z}_i(t)$ be the number of single-household epidemics that terminate before time t with a total of i recoveries and define $Z_i(t)$ similarly for the birth-death process S. Note that any such household would have been in state $(n-i, 1)$ immediately before the household epidemic ceased and hence would have finished as a type-i individual in S. Hence, recalling that the recovery rate of infectives is 1, for large t

$$Z_i(t) \approx \int_0^t Y_i(u)\mathrm{d}u \approx \int_0^t W v_i(\lambda_L, \lambda_G)\mathrm{e}^{r(\lambda_L,\lambda_G)u}\mathrm{d}u = \frac{W v_i(\lambda_L, \lambda_G)}{r(\lambda_L, \lambda_G)}(\mathrm{e}^{r(\lambda_L,\lambda_G)t} - 1).$$

Moreover, see e.g. Jagers [14], we have that

$$\mathrm{e}^{-r(\lambda_L,\lambda_G)t}Z_i(t) \xrightarrow{a.s.} W\frac{v_i(\lambda_L, \lambda_G)}{r(\lambda_G, \lambda_L)} \quad \text{as } t \to \infty. \tag{16.2}$$

Let $v_{(x,y)}(\lambda_L, \lambda_G) = v_i(\lambda_L, \lambda_G)$, where i is the label of a type-(x, y) individual in S. For $0 \leq x < n,\ y \geq 0,\ x + y \leq n$, define the function $p_{\text{multi}}(x, y|\lambda_L, \lambda_G)$ by

$$p_{\text{multi}}(x, y|\lambda_L, \lambda_G) = \begin{cases} K(\lambda_L, \lambda_G)v_{(x,y)}(\lambda_L, \lambda_G) & \text{if } y \geq 1, \\ K(\lambda_L, \lambda_G)\frac{v_{(x,1)}(\lambda_L,\lambda_G)}{r(\lambda_G,\lambda_L)} & \text{if } y = 0, \end{cases} \tag{16.3}$$

where $K(\lambda_L, \lambda_G)$ is chosen such that $\sum_{x=0}^{n-1} \sum_{y=0}^{n-x} p_{\text{multi}}(x, y|\lambda_L, \lambda_G) = 1$. Noting that only households that have been infected at some stage appear in S, (16.1) and (16.2) show that the function p_{multi} gives the asymptotic (i.e. as $t \to \infty$) proportion of households in all possible single-household epidemic states assuming the overall epidemic mimics the CMJ branching process. Assume that the early exponential growth rate r of the epidemic is observed. In practice r is estimated and there will be error in the estimation but we ignore that for the present method. Note that for fixed λ_L, the growth rate $r(\lambda_L, \lambda_G)$ is a strictly increasing function of λ_G, so if r is known, λ_L determines λ_G and we may write $\lambda_G = \lambda_G(\lambda_L, r)$. Recall that $a_{x,y}$ is the number of households in state (x, y) at the time the households epidemic is observed. Then it follows from (16.3) that λ_L can be estimated by maximising the pseudolikelihood function:

$$L_{\text{full}}(\lambda_L|\boldsymbol{a}, r) = \prod_{x=0}^{n-1} \prod_{y=0}^{n-x} p_{\text{multi}}(x, y|\lambda_L, \lambda_G(\lambda_L, r))^{a_{x,y}}.$$

Suppose that, as in the basic MPLE method, estimation is based only on completed single-household epidemics. Then λ_L may be estimated by maximising the pseudolikelihood function:

$$L_{\text{final}}(\lambda_L|\boldsymbol{a}, r) = \prod_{i=1}^{n} p_{\text{final}}(i|\lambda_L, \lambda_G(\lambda_L, r))^{a_{n-i,0}},$$

where, writing v_i for $v_i(\lambda_L, \lambda_G(\lambda_L, r))$ and using (16.3) with $y = 0$,

$$p_{\text{final}}(i|\lambda_L, \lambda_G(\lambda_L, r)) = \frac{v_i}{v_1 + v_2 + \cdots + v_n} \qquad (i = 1, 2, \cdots, n). \tag{16.4}$$

The estimator based on L_{full} assumes that the exact state of a household is observable but this may not be realised in practice. For example, it may not be possible to distinguish between susceptibles and infectives with only recoveries being observed. For $i = 1, 2, \cdots, n$, let $\mathcal{A}_i = \{(x, y) : x, y \geq 0,\ x + y = n - i\}$ be the set of household states in which there have been precisely i recoveries and let $c_i = \sum_{x=0}^{n-i} a_{x,n-i-x}$ be the observed number of households with i recoveries. Then

$$q_{\text{multi}}(i|\lambda_L, \lambda_G) = \sum_{(x,y)\in\mathcal{A}_i} p_{\text{multi}}(x, y|\lambda_L, \lambda_G) \quad (i = 1, 2, \cdots, n)$$

is the distribution of the c_i in S as $t \to \infty$. Thus, under the above assumptions, λ_L may be estimated by maximising the pseudolikelihood function:

$$L_{\text{rec}}(\lambda_L|\boldsymbol{c}, r) = \prod_{i=1}^{n} q_{\text{multi}}(i|\lambda_L, \lambda_G(\lambda_L, r))^{c_i}.$$

The asymptotic behaviour of the above estimators (derived using the asymptotic stable composition of the CMJ process) as the number of households $m \to \infty$ is investigated in Ball and Shaw [6]. For $m = 1, 2, \cdots$, let t_m denote the time when estimation is made for the epidemic amongst a population of m households. Suppose that (i) $t_m \to \infty$ as $m \to \infty$ and (ii) $t_m \leq c \log m$ for all sufficiently large m, for some $c < (2r)^{-1}$. The first condition ensures that, in the limit as $m \to \infty$, the stable composition of the CMJ branching process holds when estimation is made and the second condition ensures that the coupling of the CMJ branching process and the epidemic process still holds at that time. Recall that A_{ext} denotes the extinction set of the CMJ branching process. Suppose that a strongly consistent estimator of the growth rate r is available. Then, for P-almost all $\omega \in \Omega \setminus A_{\text{ext}}$, the above estimators converge to the true value of λ_L.

The above estimators are easily modified to allow for the rate of the exponential distribution describing the infectious period to be unknown and also estimated from the observed data. Further an exponentially distributed latent period can be included in the model. In principle, phase-type distributions (see e.g. Asmussen [1, Chap. III, Sect. 6]), can be used to permit more flexible modelling of infectious and latent periods but the state space of the corresponding multitype B-D process quickly becomes large. See Ball and Shaw [6] for further details of these, and other, extensions using the CMJ approach.

In practice, populations are finite and hence the multitype B-D process approximation of an epidemic eventually breaks down. Let $\mathcal{T}' = \{(x, y) : x \geq 0, y \geq 0, x + y < n\}$ denote the type space of all active and formerly active individuals in S. For $(x, y) \in \mathcal{T}'$, let $\tilde{a}_{x,y} = a_{x,y}/a$ be the proportion of contacted households in state (x, y) at the time of observation, where $a = \sum_{(x,y)\in\mathcal{T}'} a_{x,y}$. The TV (total variation) distance between the observed epidemic and the limiting distribution of its approximating CMJ branching process is given by

$$D(\boldsymbol{a}, \lambda_L, \lambda_G) = \sum_{(x,y)\in\mathcal{T}'} |\tilde{a}_{x,y} - p_{\text{multi}}(x, y|\lambda_L, \lambda_G)|/2.$$

Figure 16.3 shows how $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ changes as epidemics progress. Specifically epidemics with parameters $\lambda_G = 1.21$, $\lambda_L = 0.64$ and a unit-mean exponential infectious period in a population of 21,000 individuals were simulated for populations partitioned into equally-sized households of 2, 4, 6 and 8 for the left-hand plot, and populations of 5000, 10,000, 20,000 households of size 4 and a CMJ branching process made up of households of size 4 (representing an infinite population) for the right-hand plot. Each graph is based on 1000 simulated epidemics that took off. During a simulation, the distance $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ was recorded at regular timepoints based on the number of recovered individuals observed and the mean TV distance at each timepoint over the 1000 simulations provided the data points for the plots. Note that for a population of 21,000 individuals, $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ is minimised after approximately 500 recoveries have occurred regardless of the household size used. Before this point the epidemics have not had long enough in general to settle into behaviour resembling the asymptotic behaviour of the

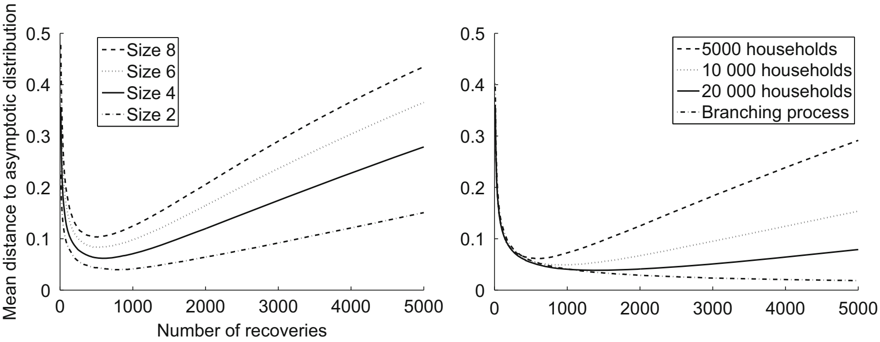


Fig. 16.3 Mean TV distances between the observed and asymptotic distribution of household states as the number of recovered individuals observed increases. See main text for further details

CMJ branching process, whilst, after this point, global infectious contacts with susceptibles in previously infected households begin to make the CMJ branching process approximation break down. It is also worth noting the general pattern of $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ increasing as household size increases. Initially this can be attributed to the smaller state space in epidemics with smaller households reducing the number of elements in the sum used to calculate $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ and allowing the epidemic to settle into its approximate CMJ branching process behaviour more quickly. As epidemics progress, the greater number of households in epidemics with smaller-sized households also means that global infectious contacts with susceptibles in previously infected households occur less frequently, so $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ remains small for longer in populations split into smaller sized households. The right-hand plot shows that as population size increases, the number of recoveries needed before the CMJ branching process approximation begins to break down also increases. For an infinite population, the mean TV distance converges towards zero as the number of recoveries increases, as predicted by theory; the mean TV distance drops quickly to about 0.05 but thereafter convergence is quite slow.

Figure 16.4 shows kernel density estimates of the distribution of the estimators of λ_L using the multitype B-D methods after 500 recoveries have occurred. The top panels use the same simulated epidemics that produced the size-2 and size-8 graphs in the left panel of Fig. 16.3, i.e. in a population of N=21,000 individuals. The bottom panels use corresponding simulations for populations with $N = 84,000$ individuals. First note that all three methods give estimates that are broadly centred around the true value of λ_L, with the full pseudolikelihood method yielding the estimator with the lowest variance since it uses the most information. When $N = 21,000$, there is a general overestimate owing to households being contacted globally more than once, but this disappears when $N = 84,000$. It is interesting to note however, that the full and recovery pseudolikelihood methods seem to be more accurate for populations split into larger household sizes, contrary to what might be expected from Fig. 16.3. This can be explained by noting that small adjustments in λ_L have a greater impact as household size increases, so larger

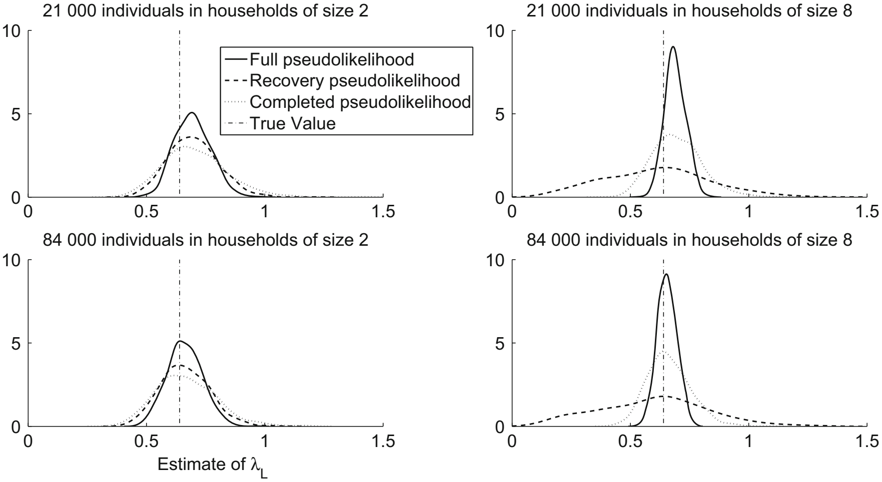


Fig. 16.4 Kernel density estimates of the distribution of the estimators of λ_L (true value 0.64 shown by *vertical lines*), using the multitype B-D methods after 500 recoveries have occurred, based on 1000 simulations that took off for each population configuration. See main text for further details

values of $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ do not necessarily result in wildly inaccurate estimates of λ_L if the TV distance $D(\boldsymbol{a}, \lambda_L, \lambda_G)$ is still relatively small. The exception to this rule is the completed households pseudolikelihood method, based on L_{final}, which appears to lose far too much information by relying on completed single-household epidemics only when household size increases to 8 and may also suffer from a lack of completed single-household epidemics when only 500 recoveries have occurred in a population split into large households.

16.4 Modified Single-Household Epidemic

In this section, still assuming that the infectious period I follows a unit-mean exponential distribution, we describe a modification of the single-household epidemic model, which we use to (i) give new derivations of both R_* and the total size distribution $p_{\text{basic}}(i|\lambda_L)$ $(i = 1, 2, \cdots, n-1)$, and (ii) prove that, if λ_L is held fixed, then, for $i = 1, 2, \cdots, n$, $q_{\text{multi}}(i|\lambda_L, \lambda_G) \to p_{\text{basic}}(i|\lambda_L)$ as $r \downarrow 0$.

The basic idea of the modified process is that whenever the single-household epidemic process terminates, it enters the graveyard state, where it remains for a time that is exponentially distributed with mean one before returning to the initial state of $n-1$ susceptibles and 1 infective. The modified process is a positive recurrent continuous-time Markov chain and the properties referred to above all follow easily from its stationary distribution.

16.4.1 Derivation of Total Size Distribution and R_*

The state space of the modified process is $\tilde{\mathcal{T}} = \mathcal{T} \cup G$, where G is the graveyard state and, as previously, $\mathcal{T} = \{(x, y) : x \geq 0, y \geq 1, x + y \leq n\}$. Give the states in $\mathcal{T}$ the labels $1, 2, \cdots, k$, precisely as in Sect. 16.3.2.1, and give the graveyard state G the label $k + 1$. Partition $\tilde{\mathcal{T}} = E \cup B \cup C \cup G$, where $E = \{1\}$ (the initial state $(n - 1, 1)$ of the single-household epidemic), $B = \{2, 3, \cdots, n\}$ (the other states with exactly 1 infective), $C = \{n + 1, n + 2, \cdots, k\}$ (the states with ≥ 2 infectives) and $G = \{k + 1\}$ (the graveyard state). Let $\{W(t)\} = \{W(t) : t \geq 0\}$ denote the modified process, using this labelling of states. Let $Q = [q_{ij}]$ be the transition-rate matrix of the continuous-time Markov chain $\{W(t)\}$. Thus Q is the $(k + 1) \times (k + 1)$ matrix with off-diagonal elements q_{ij} giving the transition rate of $\{W(t)\}$ from state i to state j and diagonal elements $q_{ii} = -\sum_{j \neq i} q_{ij}$. Partition Q into

$$Q = \begin{bmatrix} q_{EE} & \mathbf{0} & \boldsymbol{q}_{EC} & q_{EG} \\ \mathbf{0} & Q_{BB} & Q_{BC} & \boldsymbol{q}_{BG} \\ \mathbf{0} & Q_{CB} & Q_{CC} & \mathbf{0} \\ q_{GE} & \mathbf{0} & \mathbf{0} & q_{GG} \end{bmatrix},$$

where, for example, the $(n-1) \times (n-1)$ matrix Q_{BB} corresponds to transitions that remain within B (note that Q_{BB} is diagonal since transitions between distinct states in B are not possible) and the $(n-1) \times (k-n)$ matrix Q_{BC} corresponds to transitions from B to C. (Throughout the chapter, $\mathbf{0}$ denotes a (column or row) vector of zeros and $\mathbf{1}$ denotes a column vector of ones, the dimensions of which are apparent from their context.)

Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_{k+1})$ denote the stationary distribution of $\{W(t)\}$, so $\boldsymbol{\pi}$ is determined by

$$\boldsymbol{\pi} Q = \mathbf{0} \qquad \text{and} \qquad \boldsymbol{\pi}\mathbf{1} = 1.$$

Partition $\boldsymbol{\pi} = (\pi_E, \boldsymbol{\pi}_B, \boldsymbol{\pi}_C, \pi_G)$ in the obvious fashion.

The total size distribution $p_{\text{basic}}(i|\lambda_L)$ $(i = 1, 2, \cdots, n - 1)$ can be obtained from $\boldsymbol{\pi}$ as follows. When $\{W(t)\}$ leaves the graveyard state $k + 1$ it always goes to state 1. Thus, for $i = 1, 2, \cdots, n$, $p_{\text{basic}}(i|\lambda_L)$ is given by the equilibrium probability that when $\{W(t)\}$ enters G it does so from state i. Further, since $q_{i,k+1} = 1$ for $i = 1, 2, \cdots, n$, it follows that this equilibrium probability is proportional to the equilibrium probability that $\{W(t)\}$ is in state i. Hence,

$$p_{\text{basic}}(i|\lambda_L) = c\pi_i \quad (i = 1, 2, \cdots, n), \tag{16.5}$$

where $c = 1/(\pi_1 + \pi_2 + \cdots + \pi_n)$.

Turning to the threshold parameter R_*, suppose that $\{W(t)\}$ is in equilibrium and let η_G be the Markov-modulated Poisson process describing global contacts, whose rate at time t is $\lambda_G y_{W(t)}$, where y_i is the number of infectives in the state having label

i ($i = 1, 2, \cdots, k$) and $y_{k+1} = 0$. (Thus if $(x, y) \in \mathcal{T}$ has label i then $y_i = y$.) Then η_G has intensity given by

$$\lambda_G \left(\pi_E y_E + \boldsymbol{\pi}_B \boldsymbol{y}_B + \boldsymbol{\pi}_C \boldsymbol{y}_C\right),$$

where, for example, $\boldsymbol{y}_B = (y_j : j \in B)^\top = (y_2, y_3, \cdots, y_n)^\top$, with $^\top$ denoting transpose. Let η_E be the point process describing the times when $\{W(t)\}$ enters state 1. The intensity of η_E is $\pi_{k+1} q_{k+1,1} = \pi_G$, since $q_{k+1,1} = 1$, so the mean time between two successive entries of $\{W(t)\}$ to state 1 is π_G^{-1}. Now R_* is the mean number of points of η_G between two successive entries of $\{W(t)\}$ to state 1, so

$$R_* = \pi_G^{-1} \lambda_G \left(\pi_E y_E + \boldsymbol{\pi}_B \boldsymbol{y}_B + \boldsymbol{\pi}_C \boldsymbol{y}_C\right). \tag{16.6}$$

16.4.2 Critical Epidemics, i.e. $r = 0$

The matrix $A = \Lambda - \text{diag}(\boldsymbol{\mu})$, see Sect. 16.3.2.1, is given in partitioned form by

$$A = \begin{bmatrix} \lambda_G y_E + q_{EE} & \boldsymbol{0} & \boldsymbol{q}_{EC} \\ \lambda_G \boldsymbol{y}_B & Q_{BB} & Q_{BC} \\ \lambda_G \boldsymbol{y}_C & Q_{CB} & Q_{CC} \end{bmatrix}.$$

Recall that $\boldsymbol{v}(\lambda_L, \lambda_G)$ is the left-eigenvector of A associated with the eigenvalue $r(\lambda_L, \lambda_G)$, normalised so that $\boldsymbol{v}(\lambda_L, \lambda_G)\boldsymbol{1} = 1$. For ease of notation, write $\boldsymbol{v}$ for $\boldsymbol{v}(\lambda_L, \lambda_G)$ and r for $r(\lambda_L, \lambda_G)$.

Lemma 16.1 *When $r = 0$, the corresponding left-eigenvector $\boldsymbol{v}$ of A is given by*

$$\boldsymbol{v} = (\pi_E, \boldsymbol{\pi}_B, \boldsymbol{\pi}_C)/(\pi_E + \boldsymbol{\pi}_B \boldsymbol{1} + \boldsymbol{\pi}_C \boldsymbol{1}).$$

Proof Expanding $\boldsymbol{\pi} Q = \boldsymbol{0}$ in partitioned form yields

$$(\pi_E, \boldsymbol{\pi}_B, \boldsymbol{\pi}_C) \begin{bmatrix} \boldsymbol{0} & \boldsymbol{q}_{EC} \\ Q_{BB} & Q_{BC} \\ Q_{CB} & Q_{CC} \end{bmatrix} = 0 \qquad \text{and} \qquad \pi_E q_{EE} + \pi_G q_{GE} = 0. \tag{16.7}$$

Suppose that $\boldsymbol{v} = (\pi_E, \boldsymbol{\pi}_B, \boldsymbol{\pi}_C)$ and $r = 0$. Then, using the first equation in (16.7), $\boldsymbol{v}A = \boldsymbol{0}$ if and only if

$$\pi_E q_{EE} + \lambda_G (\pi_E y_E + \boldsymbol{\pi}_B \boldsymbol{y}_B + \boldsymbol{\pi}_C \boldsymbol{y}_C) = 0. \tag{16.8}$$

The second equation in (16.7) implies that $\pi_E q_{EE} = -\pi_G$ (recall that $q_{GE} = 1$). Also, $R_* = 1$ when $r = 0$, so (16.6) gives $\pi_G = \lambda_G (\pi_E y_E + \boldsymbol{\pi}_B \boldsymbol{y}_B + \boldsymbol{\pi}_C \boldsymbol{y}_C)$. Thus (16.8) is satisfied and $\boldsymbol{v}A = \boldsymbol{0}$, as required. □

Note that (16.4), (16.5) and Lemma 16.1 imply that, for $i = 1, 2, \cdots, n$,

$$p_{\text{final}}(i|\lambda_L, \lambda_G(\lambda_L, r)) \to p_{\text{basic}}(i|\lambda_L) \quad \text{as } r \downarrow 0,$$

so the basic method yields an approximately unbiased estimator if r is small.

16.5 Concluding Comments

When fitting household and other models to data on an emerging epidemic, the data collected need to be modelled very carefully taking due account of the emerging nature of the epidemic. We have demonstrated that the asymptotic theory of continuous-time, multitype Markov branching processes provides a feasible approach to modelling such data for Markov epidemics. The stable composition of supercritical branching processes provides, in principle, a more flexible approach, as it is not restricted to Markov models, but its implementation for non-Markov models is generally difficult and an area for future research. Other important areas for further research include developing approximate confidence intervals for the estimates and extending the methodology to (i) multitype epidemics (for example, to age-stratified populations with age-specific transmission rates), (ii) allow for asymptomatic cases and (iii) incorporate information on temporal progression of disease within households.

Acknowledgements Laurence Shaw was supported by an EPSRC Doctoral Training grant.

References

1. Asmussen, S.: Applied Probability and Queues. Wiley, New York (1987)
2. Athreya, K.E., Ney, P.E.: Branching Processes. Springer, Berlin (1972)
3. Ball, F.: A unified approch to the distribution of total size and total area under the trajectory of infectives in epidemic models. Adv. Appl. Probab. **18**, 289–310 (1986)
4. Ball, F.: Threshold behaviour in stochastic epidemics among households. In: Athens Conference on Applied Probability and Time Series Analysis. Vol. I. Lecture Notes in Statistics, vol. 114, pp. 253–266. Springer, Berlin (1996)
5. Ball, F., Donnelly, P.: Strong approximations for epidemic models. Stoch. Proc. Appl. **55**, 1–21 (1995)
6. Ball, F., Shaw, L.: Estimating the within-household infection rate in emerging SIR epidemics among a community of households. J. Math. Biol. **71**, 1705–1735 (2015)
7. Ball, F., Mollison, D., Scalia-Tomba, G.: Epidemics with two levels of mixing. Ann. Appl. Probab. **7**, 46–89 (1997)
8. Bartlett, M.S.: An Introduction to Stochastic Processes. Cambridge University Press, Cambridge (1955)
9. Becker, N.G., Dietz, K.: The effect of the household distribution on transmission and control of highly infectious diseases. Math. Biosci. **127**, 207–219 (1995)

10. Cauchemez, S., Carrat, F., Viboud, C., Valleron, A.J., Boëlle, P.Y.: A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Stat. Med. **23**, 3469–3487 (2004)
11. Fraser, C.: Estimating individual and household reproduction numbers in an emerging epidemic. PLoS ONE **2**(8), e758 (2007)
12. Heesterbeek, J.A.P., Dietz, K.: The concept of R_0 in epidemic theory. Statistica Neerlandica **50**, 89–110 (1996)
13. House, T., Inglis, N., Ross, J.V., Wilson, F., Suleman, S., Edeghere, O., Smith, G., Olowokure, B., Keeling, M.J.: Estimation of outbreak severity and transmissibility: Influenza A(H1N1)pdm09 in households. BMC Med. **10**, 117 (2012)
14. Jagers, P.: Stabilities and instabilities in population dynamics. J. Appl. Probab. **29**, 770–780 (1992)
15. Kendall, D.G.: Deterministic and stochastic epidemics in closed populations. In: Proceedings of 3rd Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 149–165. University of California Press, Berkeley (1956)
16. Nerman, O.: On the convergence of supercritical gerneral (C-M-J) branching processes. Z. Wahrscheinlichkeitstheorie **57**, 365–395 (1981)
17. Wallinga, J., Lipsitch, M.: How generation intervals shape the relationship between growth rates and reproductive numbers. Proc. R. Soc. Lond. B **274**, 599–604 (2007)
18. Whittle, P.: The outcome of a stochastic epidemic — a note on Bailey's paper. Biometrika **42**, 116–122 (1955)

Part VIII
Applications in Biology and Genetics

Chapter 17
Extinction Probability of Some Recessive Alleles of X-Linked Genes in the Context of Two-Sex Branching Processes

Miguel González, Cristina Gutiérrez, Rodrigo Martínez, and Manuel Mota

Mathematics Subject Classification (2000): 60J80, 60J85

17.1 Introduction

It is well-known that the responsible of the sex-determination in humans and some animal populations are a pair of chromosomes, X and Y. Females carry XX chromosomes, while males have XY chromosomes. Certain characteristics or diseases are due to genes linked to X chromosome, to Y chromosome or to both sexual chromosomes. If a gene has a pair of alleles R and r, females and males of various genotypes and/or phenotypes appear in the population. For example, a gene linked to Y chromosome is only associated to males and two genotypes are possible depending on allele (R or r) the male carries. Recently, some models describing the evolution of the two genotypes defined by a Y-linked gene have been investigated (see for example [1, 5–10]).

If the gene is linked to X chromosome, females and males of various genotypes and/or phenotypes appear in the population. Specifically, we can find three genotypes for females:

$$X^RX^R, \quad X^RX^r, \quad X^rX^r;$$

M. González (✉) • C. Gutiérrez • M. Mota
Department of Mathematics, University of Extremadura, Avda. Elvas, 06006 Badajoz, Spain
e-mail: mvelasco@unex.es; cgutierrez@unex.es; mota@unex.es

R. Martínez
Centro Universitario de Plasencia, Department of Mathematics, University of Extremadura, Plasencia, Spain
e-mail: rmartinez@unex.es

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_17

and two genotypes for males

$$X^R Y, \quad X^r Y.$$

Notice that, as usual in Genetics, genotype $X^R X^r$ also represents the combination $X^r X^R$.

If the allele R is dominant and the allele r is recessive two phenotypes are possible for both, females and males: R phenotype (RR and Rr females and R males) and r phenotype (rr females and r males). A general model to describe the evolution of the number of carriers of the genotypes and phenotypes defined by the alleles of an X-linked gene becomes very complex from a theoretical point of view.

An interesting situation occurs when one of the alleles is responsible for a serious disorder or for a severe disease. If the dominant allele (R) is defective, all the carriers are affected and most of them do not reach breeding age so it is rarely detected due to its rapid elimination from populations. However, if the pernicious allele is recessive (r), individuals with r phenotype do not last in the population, but heterozygous carrier females are able to live and reproduce. They do not phenotypically express the genetic condition but can pass r allele onto offspring. As example of such genes one can find the responsible for hemophilia, Duchenne muscular dystrophy or other genetic diseases (see [4] and references therein).

In this work we propose a branching model to describe the evolution of the number of carriers of such recessive pernicious alleles in a population. Since the r phenotype individuals, affected by a serious disorder, are hardly selected as mates, it seems realistic to consider that only R phenotype individuals participate in the mating phase. Therefore, in the proposed model two types of couples can be formed, according to the genotype of the female: $RR \times R$ *couples* (homozygous female) and $Rr \times R$ *couples* (heterozygous females), which will be referred as the *genotypes of the couples*. Notice that as consequence of this assumption, rr females are not taken into account in the model whereas r males are considered among the offspring of an $Rr \times R$ couple but without participation in the subsequent mating.

We provide the definition of the model in Sect. 17.2. In Sect. 17.3, we study the moments of the underlying random variables of the model and provide a classification of states of the defined process. Section 17.4 is devoted to study the fate of the recessive allele in the population, showing its extinction-explosion dichotomy and providing conditions under which it eventually disappears. After that, we investigate the fixation of the dominant allele in Sect. 17.5. In Sect. 17.6 we study the long term coexistence of both alleles in the population through some conjectures that have been checked via simulation. The proofs of these conjectures are open problems that can be viewed as a possible way to continue the investigations. Finally, the proofs of the statements are relegated to Sect. 17.7 for a better reading of the chapter.

17.2 The Model

In this section we provide the formal definition of the model. Previously, we give the intuitive meaning of the different mathematical objects involved in the definition.

Mathematically, the model we introduce is a multitype two-sex branching model. Thus, the modeling of such populations includes two phases: the reproduction and the mating. Let us formulate the assumptions we make in relation to both phases, providing also their intuitive interpretation.

17.2.1 Working Hypotheses: Reproduction

As pointed out above, we consider only two types of couples: $RR \times R$ and $Rr \times R$ couples. At time of the reproduction we suppose that the total offspring of an $RR \times R$ couple is a random variable on the non-negative integers with probability distribution $\{p_k^R\}_{k\geq 0}$. Analogously the total offspring of an $Rr \times R$ couple has probability distribution $\{p_k^r\}_{k\geq 0}$. These distributions are called reproduction laws of the $RR \times R$ and $Rr \times R$ couples, respectively, and we denote by m_R and m_r their respective means. We assume that the variances of these distributions are both finite.

With respect to the sex and the genotype of an offspring, in an $RR \times R$ couple the female transmits an X^R chromosome with probability 1. If we assume that a male gives his X^R chromosome with probability α $(0 < \alpha < 1)$ and his Y chromosome with probability $1-\alpha$, then an $RR \times R$ couple offspring is RR female with probability α and R male with probability $1-\alpha$.

In an $Rr \times R$ couple, this issue becomes a bit more complex. While the behaviour assumed for the males is the same as described above, now an Rr female is supposed to give her X^R chromosome with probability β $(0 < \beta < 1)$ and her X^r chromosome with probability $1-\beta$. Hence, an offspring of an $Rr \times R$ couple is an RR female with probability $\alpha\beta$, an Rr female with probability $\alpha(1-\beta)$, an R male with probability $(1-\alpha)\beta$ and an r male with probability $(1-\alpha)(1-\beta)$.

17.2.2 Working Hypotheses: Mating

We assume that every individual mates with only one individual of the opposite sex, providing there is some of these individuals without a mate, i.e. a perfect fidelity scheme (see e.g. [3]). Also, as pointed out above, we suppose rr females are not taken into account and r males do not mate, i.e. RR and Rr females mate only with R males. According this scheme, if the number of R males exceeds that of females, the number of $RR \times R$ and $Rr \times R$ couples is, respectively, the same as the number of RR and Rr females. On the other hand, if there are less R males than females, all these males mate. Since the defective allele does not appear in the phenotype of Rr

females, the R males make a *blind choice* of their mates, which is modeled through the Hypergeometric distribution.

17.2.3 Mathematical Definition of the Model

In order to provide a formal definition of a model with such working hypotheses, let us consider the following independent sequences of independent and identically distributed, non-negative and integer valued random vectors:

$$\{(F_{nl}^{RR,RR\times R}, M_{nl}^{R,RR\times R}) : l = 1,2,\dots ; n = 0,1,\dots\} \quad \text{and}$$
$$\{(F_{nl}^{RR,Rr\times R}, F_{nl}^{Rr,Rr\times R}, M_{nl}^{R,Rr\times R}, M_{nl}^{r,Rr\times R}) : l = 1,2,\dots ; n = 0,1,\dots\}.$$

- $(F_{nl}^{RR,RR\times R}, M_{nl}^{R,RR\times R})$ represents the number of RR females and R males generated by the lth $RR \times R$ couple of the nth generation.
- $(F_{nl}^{RR,Rr\times R}, F_{nl}^{Rr,Rr\times R}, M_{nl}^{R,Rr\times R}, M_{nl}^{r,Rr\times R})$ represents the number of RR females, Rr females, R males and r males generated by the lth $Rr \times R$ couple of the nth generation.

The common distribution of the random variables $F_{nl}^{RR,RR\times R} + M_{nl}^{R,RR\times R}$ (resp. $F_{nl}^{RR,Rr\times R} + F_{nl}^{Rr,Rr\times R} + M_{nl}^{R,Rr\times R} + M_{nl}^{r,Rr\times R}$) is $\{p_k^R\}_{k\geq 0}$ (resp. $\{p_k^r\}_{k\geq 0}$).

Moreover, the joint probability distribution of $(F_{nl}^{RR,RR\times R}, M_{nl}^{R,RR\times R})$ given that $F_{nl}^{RR,RR\times R} + M_{nl}^{R,RR\times R} = k$ is multinomial with parameters k, α, $1-\alpha$. Analogously, the joint probability distribution of the vector $(F_{nl}^{RR,Rr\times R}, F_{nl}^{Rr,Rr\times R}, M_{nl}^{R,Rr\times R}, M_{nl}^{r,Rr\times R})$ given that the sum of its components is equal to k is multinomial with parameters k, $\alpha\beta$, $\alpha(1-\beta)$, $(1-\alpha)\beta$, $(1-\alpha)(1-\beta)$.

In what follows, for every $n \geq 0$, $Z_n^{RR\times R}$ and $Z_n^{Rr\times R}$ denote the total number of $RR \times R$ and $Rr \times R$ couples, respectively, at generation n. Also, F_n^{RR}, F_n^{Rr}, M_n^R and M_n^r denote the numbers of RR females, Rr females, R males and r males, respectively, at generation n.

We define the *X-linked two-sex branching process*, $\{(F_n^{RR}, F_n^{Rr}, M_n^R, M_n^r)\}_{n\geq 1}$, and the associated sequence of mating units $\{(Z_n^{RR\times R}, Z_n^{Rr\times R})\}_{n\geq 0}$, recursively as follows:

- If, at generation n, there are $Z_n^{RR\times R}$ couples of type $RR \times R$ and $Z_n^{Rr\times R}$ couples of type $Rr \times R$, then:
    - The total offspring of the $RR \times R$ couples is

$$(F_{n+1}^{RR,RR\times R}, M_{n+1}^{R,RR\times R}) = \sum_{l=1}^{Z_n^{RR\times R}} (F_{nl}^{RR,RR\times R}, M_{nl}^{R,RR\times R}).$$

- The total offspring of the $Rr \times R$ couples is

$$(F_{n+1}^{RR,Rr\times R}, F_{n+1}^{Rr,Rr\times R}, M_{n+1}^{R,Rr\times R}, M_{n+1}^{r,Rr\times R}) = \sum_{l=1}^{Z_n^{Rr\times R}} (F_{nl}^{RR,Rr\times R}, F_{nl}^{Rr,Rr\times R}, M_{nl}^{R,Rr\times R}, M_{nl}^{r,Rr\times R}).$$

The number of RR females, Rr females, R males and r males at generation $n+1$ is, respectively:

$$F_{n+1}^{RR} = F_{n+1}^{RR,RR\times R} + F_{n+1}^{RR,Rr\times R}, \quad F_{n+1}^{Rr} = F_{n+1}^{Rr,Rr\times R},$$

$$M_{n+1}^{R} = M_{n+1}^{R,RR\times R} + M_{n+1}^{R,Rr\times R}, \qquad M_{n+1}^{r} = M_{n+1}^{r,Rr\times R},$$

and therefore, the total number of females and males in the $(n+1)$st generation is, respectively:

$$F_{n+1} = F_{n+1}^{RR} + F_{n+1}^{Rr}, \quad M_{n+1} = M_{n+1}^{R} + M_{n+1}^{r}.$$

- Since monogamous mating is assumed and r males do not participate in the mating phase, the number of couples formed by F_{n+1} females and M_{n+1}^R males at generation $n+1$ is

$$Z_{n+1} = Z_{n+1}^{RR\times R} + Z_{n+1}^{Rr\times R} = \min\{F_{n+1}, M_{n+1}^R\}.$$

Furthermore, if $F_{n+1} \leq M_{n+1}^R$, i.e. there are less females than R males, then all the females mate and therefore the number of couples of each genotype is equal to the number of females of that genotype:

$$Z_{n+1}^{RR\times R} = F_{n+1}^{RR}, \qquad Z_{n+1}^{Rr\times R} = F_{n+1}^{Rr}.$$

Finally, if $F_{n+1} > M_{n+1}^R$, i.e. there are more females than R males, then all these males mate, and consequently they must choose blindly the genotype of their mates. This blind choice is modeled through the Hypergeometric distribution:

$$Z_{n+1}^{RR\times R} | (F_{n+1}^{RR}, F_{n+1}^{Rr}, M_{n+1}^{R}, M_{n+1}^{r}) \sim Hypergeometric(M_{n+1}^{R}, F_{n+1}, F_{n+1}^{RR}),$$

$$Z_{n+1}^{Rr\times R} = M_{n+1}^{R} - Z_{n+1}^{RR\times R}.$$

17.3 Basic Properties of the Model

In this section, we study first some properties of the conditional moments of the variables of the model. Some of these properties are used to prove asymptotic results related to the extinction/survival of one of the alleles or of the population.

The following results can be deduced as an immediate consequence of the definition of the model.

Proposition 17.1 *The following equalities hold almost surely (a.s.):*

(i) $E[(Z_n^{RR\times R}, Z_n^{Rr\times R})|(F_n^{RR}, F_n^{Rr}, M_n^R, M_n^r)] = (F_n^{RR}, F_n^{Rr})\min\{1, F_n^{-1}M_n^R\}$.
(ii) $E[F_{n+1}^{RR}|(Z_n^{RR\times R}, Z_n^{Rr\times R})] = Z_n^{RR\times R}\alpha m_R + Z_n^{Rr\times R}\alpha\beta m_r$.
(iii) $E[F_{n+1}^{Rr}|(Z_n^{RR\times R}, Z_n^{Rr\times R})] = Z_n^{Rr\times R}\alpha(1-\beta)m_r$.
(iv) $E[M_{n+1}^{R}|(Z_n^{RR\times R}, Z_n^{Rr\times R})] = Z_n^{RR\times R}(1-\alpha)m_R + Z_n^{Rr\times R}(1-\alpha)\beta m_r$.
(v) $E[M_{n+1}^{r}|(Z_n^{RR\times R}, Z_n^{Rr\times R})] = Z_n^{Rr\times R}(1-\alpha)(1-\beta)m_r$.

From the definition it is easily deduced that $\{(F_n^{RR}, F_n^{Rr}, M_n^R, M_n^r)\}_{n\geq 1}$ and $\{(Z_n^{RR\times R}, Z_n^{Rr\times R})\}_{n\geq 0}$, are homogeneous Markov chains. Our next objective is to establish some properties involving the states of $\{(Z_n^{RR\times R}, Z_n^{Rr\times R})\}_{n\geq 0}$ and whose proof (which is omitted) is obtained by taking into account the multinomial scheme of the reproduction laws and applying a standard procedure.

Proposition 17.2

(i) $(0,0)$ *is an absorbing state.*
(ii) *Every non-null state* $(i,j) \neq (0,0)$ *is transient.*
(iii) *If* $p_0^R + p_1^R + p_2^R + p_3^R < 1$ *and* $p_0^r + p_1^r + p_2^r + p_3^r < 1$, *then the sets* $\{(i,j), i \geq 0, j > 0\}$ *and* $\{(i,0), i > 0\}$ *are classes of communicating states and each state leads to the state* $(0,0)$. *Furthermore, the states belonging to the first set may move to the other in one step.*

Remark 17.1 From Proposition 17.2, it seems clear that the behaviour of the r allele in the population is not the same as the behaviour of the R allele. In fact, it is immediate to see that if in a generation there are no $Rr\times R$ couples in the population, then r allele disappears from the population, that is

$$\{Z_{n_0}^{Rr\times R} = 0\} \subseteq \{Z_{n_0+n}^{Rr\times R} = 0\}, \qquad \text{for every } n > 0.$$

In this scenario, fixation of R allele occurs. In fact, the process $\{Z_{n_0+n}^{RR\times R}\}_{n\geq 0}$ is a Bisexual Galton-Watson Branching Process (BBP) with perfect fidelity mating function and reproduction law $\{p_k^R\}_{k\geq 0}$ (see [3]).

On the contrary, RR females and R males can be produced by $RR\times R$ and $Rr\times R$ couples, so the R allele only disappears if the population becomes extinct.

According to previous remark, in our subsequent investigation it seems logical to deal first with the recessive allele, r, and then to study the dominant allele, R.

Henceforth, to simplify the notation, we denote $P(\cdot | Z_0^{RR\times R} = i, Z_0^{Rr\times R} = j)$ by $P_{(i,j)}(\cdot)$. Even (i, j) is dropped from this notation if there is no ambiguity.

17.4 The Fate of the Recessive Allele in the Population

In this section we focus on studying the behaviour of the number of carriers of the r allele. This study is made through the process $\{Z_n^{Rr\times R}\}_{n\geq 0}$, since $F_k^{Rr} = 0$ and $M_k^r = 0$, for all $k > n$ if $Z_n^{Rr\times R} = 0$. Although that process is not a homogeneous Markov chain, we establish in the following result that it shows the dual asymptotic extinction-explosion behaviour, typical of many homogeneous branching processes: the total number of couples of that genotype either goes to zero or undergoes unlimited growth. Intuitively, this duality must be interpreted as the extinction or persistence of the recessive allele.

According to this interpretation, the event $\{Z_n^{Rr\times R} \to 0\}$ is called *extinction of the r allele*, and the event $\{Z_n^{Rr\times R} \to \infty\}$ is called *survival of the r allele*.

Theorem 17.1 *It is true that* $P(Z_n^{Rr\times R} \to 0) + P(Z_n^{Rr\times R} \to \infty) = 1$.

In the following result, we show conditions for the event *extinction of the r allele* to occur with probability one. These conditions are given in terms of the parameters $\alpha(1-\beta)m_r$ and $(1-\alpha)(1-\beta)m_r$. Intuitively, they correspond to the mean number of females and males, respectively, with r allele generated by an $Rr \times R$-couple.

Theorem 17.2 *Let* $i \geq 0, j > 0$. *Then* $P_{(i,j)}(Z_n^{Rr\times R} \to 0) = 1$ *if at least one of the following conditions is verified:*

(i) $\alpha(1-\beta)m_r \leq 1$,
(ii) $(1-\alpha)(1-\beta)m_r < 1$.

Remark 17.2 The case $(1-\alpha)(1-\beta)m_r = 1$ is not considered explicitly in Theorem 17.2. However, if $\alpha \leq 0.5$ and $(1-\alpha)(1-\beta)m_r = 1$, then $\alpha(1-\beta)m_r \leq 1$, and Theorem 17.2 (i) guarantees the extinction of the r allele. Therefore, it is the case $\alpha > 0.5$ and $(1-\alpha)(1-\beta)m_r = 1$ which is not covered by the result.

17.5 Fixation of the Dominant Allele in the Population

In this section we investigate the evolution of the number of carriers of the R allele in the population when the r allele becomes extinct. Notice that R allele is present in RR and Rr females, and R males, that are produced by both, $RR \times R$ and $Rr \times R$ couples. If r allele is extinct only $RR \times R$ couples can produce individuals, so that the fixation R allele (that is, its survival) depends only on this type of couples. This implies that the study of the fixation of the dominant allele can be developed through the process $\{Z_n^{RR\times R}\}_{n\geq 0}$.

However, as was pointed out in Remark 17.1, on the event $\{Z_n^{Rr\times R} \to 0\}$, the process $\{Z_n^{RR\times R}\}_{n\geq 0}$ evolves as a BBP with perfect fidelity mating and reproduction law $\{p_k^R\}_{k\geq 0}$, at least from one generation on (possibly different for each path). Therefore, this process also presents the dual extinction-explosion behaviour on the event *extinction of the r allele*, and the following result is established:

Theorem 17.3 *It is verified that, almost surely,*

$$\{Z_n^{Rr\times R} \to 0\} = \{Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to \infty\} \cup \{Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to 0\}.$$

Next we focus on each one of the events presented in the above result. To this end, we use the term *fixation of R allele* for the set $\{Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to \infty\}$, and *extinction of the population* for the set $\{Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to 0\}$. We study conditions for the first to occur with positive probability and for the second to have a probability of one or less than one.

As was noted above, in this case the R allele behaves, from some generation on, as a BBP with perfect fidelity mating. Hence, the theory developed in [3] can be applied here, and one then immediately deduces the following result:

Theorem 17.4 *Let* $i \geq 0, j > 0$. *Then* $P_{(i,j)}(Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to \infty) > 0$ *if and only if* $\min\{\alpha m_R, (1-\alpha)m_R\} > 1$.

Intuitively, this result states that a necessary and sufficient condition for the R allele to have a positive probability of fixation is that both the female and the male offspring per $RR \times R$ couple are on average greater than one, independently of the values that take the parameters of the reproduction law of $Rr \times R$ couples. This is due to the fact that the event $\{Z_n^{Rr\times R} \to 0\}$ always has positive probability.

According to the definition of *extinction of the population*, this event, mathematically expressed as $\{Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to 0\}$, occurs if from some generation on there are no couples of any type. Now, we investigate under which conditions this event occurs with probability one or less than one.

From Theorem 17.2 we deduce that if the average number of Rr females is less than or equal to one or the mean number of r males stemming from an $Rr\times R$ couple is less than one, then the r allele becomes extinct, so the extinction of the population is equivalent to the extinction of $RR \times R$ couples. If it is also satisfied that the mean number of females and males stemming from an $RR\times R$ couple is less than or equal to one, Theorem 17.4 ensures that fixation of the R allele is impossible and from Theorem 17.3 one deduces that the population becomes extinct almost surely.

There exists, however, a positive probability of survival of the population when the mean numbers of females and males generated by an $RR \times R$ couple are greater than one (see Theorem 17.4). Summarizing, we establish the following result:

Corollary 17.1 *Let* $i \geq 0, j > 0$. *It is verified that*

(i) *If* $\alpha(1-\beta)m_r \leq 1$ *or* $(1-\alpha)(1-\beta)m_r < 1$ *and* $\min\{\alpha m_R, (1-\alpha)m_R\} \leq 1$ *then* $P_{(i,j)}(Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to 0) = 1$.

(ii) *If* $\min\{\alpha m_R, (1-\alpha)m_R\} > 1$ *then* $P_{(i,j)}(Z_n^{Rr\times R} \to 0, Z_n^{RR\times R} \to 0) < 1$.

17.6 Coexistence of Both Alleles in the Population

In this section we deal with the problem of finding conditions that guarantee a positive probability of coexistence of both alleles in the population. Taking into account Theorem 17.1, we understand by coexistence the event $\{Z_n^{RR\times R} \not\to 0, Z_n^{Rr\times R} \to \infty\}$. Since only $Rr\times R$ couples produce individuals carrying both alleles, coexistence is equivalent to the event $\{Z_n^{Rr\times R} \to \infty\}$. Considering the results of Sect. 17.4 we assume that $\min\{\alpha(1-\beta)m_r, (1-\alpha)(1-\beta)m_r\} > 1$. In this scenario we provide some conjectures that have been derived from reasonings based on expected values and long term behaviour of the sequences involved in the X-linked two-sex branching process and their validity have been checked computationally, through simulated examples. In our discussion we differentiate the cases $\alpha \geq 0.5$ and $\alpha < 0.5$, though some results seem to hold in both situations.

17.6.1 *Case* $\alpha \geq 0.5$

If $\alpha \geq 0.5$, the couples produce on average less males than females, which eventually should balance the population towards females. In consequence, one could expect that $M_n^R \leq F_n$ for n large. Therefore, the number of couples of each type would be determined by hypergeometric distribution which, jointly with the conditioned moments calculated in Proposition 17.1, gives the following approximations for large n:

$$Z_{n+1}^{RR\times R} \simeq (F_{n+1}^{RR,RR\times R} + F_{n+1}^{RR,Rr\times R})\frac{M_{n+1}^R}{F_{n+1}} \simeq (\alpha m_R Z_n^{RR\times R} + \alpha\beta m_r Z_n^{Rr\times R})\frac{M_{n+1}^R}{F_{n+1}}, \tag{17.1}$$

$$Z_{n+1}^{Rr\times R} \simeq F_{n+1}^{Rr,Rr\times R}\frac{M_{n+1}^R}{F_{n+1}} \simeq \alpha(1-\beta)m_r Z_n^{Rr\times R}\frac{M_{n+1}^R}{F_{n+1}}. \tag{17.2}$$

Provided $Z_{n+1}^{Rr\times R} \neq 0$, we can calculate the ratio between (17.1) and (17.2):

$$\frac{Z_{n+1}^{RR\times R}}{Z_{n+1}^{Rr\times R}} \simeq \frac{m_R}{(1-\beta)m_r}\frac{Z_n^{RR\times R}}{Z_n^{Rr\times R}} + \frac{\beta}{1-\beta}. \tag{17.3}$$

Taking limits as n tends to infinity on both sides of this expression, we can derive our first conjecture:

Conjecture 17.1 On $\{Z_n^{Rr\times R} \to \infty\}$, the sequence $\{Z_n^{RR\times R}/Z_n^{Rr\times R}\}_{n=0}^{\infty}$ converges almost surely to a constant γ. Moreover, $\gamma = ((1-\beta)m_r - m_R)^{-1}\beta m_r$ if $(1-\beta)m_r > m_R$ and $\gamma = \infty$ if $(1-\beta)m_r \leq m_R$.

Intuitively, we claim that the growth of the numbers of $RR \times R$ couples is, at least, of the same order as the number of $Rr \times R$ couples. Hence, as pointed out above, the persistence of the r allele in the population is equivalent to the coexistence, and also $\{Z_n^{Rr\times R} \to \infty\} = \{Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty\}$.

Using this conjecture, we can also find out the rates of growth of the sequences $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$, and therefore establish conditions under which coexistence holds. In fact, if we take again into account Proposition 17.1, it seems reasonable to assume that

$$\frac{M_{n+1}^R}{F_{n+1}} \simeq \frac{1-\alpha}{\alpha} \, \frac{\beta m_r Z_n^{Rr\times R} + m_R Z_n^{RR\times R}}{m_r Z_n^{Rr\times R} + m_R Z_n^{RR\times R}}. \tag{17.4}$$

In the case $(1-\beta)m_r > m_R$, $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ have the same rate of growth and the parameter $\gamma < \infty$ indicates the (odds) ratio between the numbers of $RR\times R$ and $Rr\times R$ couples in the long term. Consequently, taking into account (17.2) and (17.4), it should happen that

$$Z_{n+1}^{Rr\times R} \simeq \tau Z_n^{Rr\times R} \quad \text{with} \quad \tau = \left(\frac{\beta m_r + \gamma m_R}{m_r + \gamma m_R}\right)(1-\alpha)(1-\beta)m_r.$$

In the case $(1-\beta)m_r < m_R$, both sequences grow geometrically with different rates of growth and, in particular, $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ is negligible with respect to $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$, so the quotient in the right-hand side of (17.4) converges to $(1-\alpha)/\alpha$ and therefore, from (17.1) and (17.2) we can derive that

$$Z_{n+1}^{RR\times R} \simeq (1-\alpha)m_R Z_n^{RR\times R} + (1-\alpha)\beta m_r Z_n^{Rr\times R}$$

and

$$Z_{n+1}^{Rr\times R} \simeq (1-\alpha)(1-\beta)m_r Z_n^{Rr\times R}.$$

Special attention deserves the case $(1-\beta)m_r = m_R$, where from (17.3) one deduces that the quotient between the rates of growth of both sequences is of a linear order, that is

$$Z_{n+1}^{RR\times R} \simeq (n+1)\frac{\beta}{1-\beta} Z_{n+1}^{Rr\times R},$$

so the rates of growth that one could expect for $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ are $((1-\alpha)(1-\beta)m_r)^n$ and $n((1-\alpha)(1-\beta)m_r)^{n-1}$, respectively.

The following conjecture summarizes the previous reasoning:

Conjecture 17.2 Assume $\alpha \geq 0.5$ and $\min\{\alpha(1-\beta)m_r, (1-\alpha)(1-\beta)m_r\} > 1$, and let $i \geq 0, j > 0$.

(i) If $(1-\beta)m_r > m_R$, then $P_{(i,j)}(Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty) > 0$ if and only if $\tau > 1$. Moreover, if $\tau > 1$, $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ grow at a rate of τ^n.

(ii) If $(1-\beta)m_r < m_R$, then $P_{(i,j)}(Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty) > 0$. Moreover, $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ grows at a rate of $((1-\alpha)m_R)^n$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ grows at a rate of $((1-\alpha)(1-\beta)m_r)^n$.

(iii) If $(1-\beta)m_r = m_R$, then $P_{(i,j)}(Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty) > 0$. Moreover, $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ grows at a rate of $((1-\alpha)(1-\beta)m_r)^n$ and $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ grows at a rate of $n((1-\alpha)(1-\beta)m_r)^{n-1}$.

Remark 17.3 Notice that, according to Conjecture 17.2(i), if $\tau \leq 1$ and $(1-\beta)m_r > m_R$, then coexistence has null probability or equivalently $P(Z_n^{Rr\times R} \to 0) = 1$, despite $\min\{\alpha(1-\beta)m_r, (1-\alpha)(1-\beta)m_r\} > 1$. Moreover, applying Theorem 17.4, either extinction or fixation of R allele occurs depending on whether $\min\{\alpha m_R, (1-\alpha)m_R\} = (1-\alpha)m_R \leq 1$ or > 1. In the last case, the rate of growth of $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ is $((1-\alpha)m_R)^n$.

Next we show a simulated example that illustrate Conjecture 17.2(iii). We have simulated an X-linked two-sex branching process with $\alpha = 0.5$, $\beta = 0.7$, $m_r = 8$ and, consequently, $m_R = (1-\beta)m_r = 2.4$ so Conjecture 17.1(i) is satisfied with $\gamma = \infty$. Moreover, $(1-\alpha)(1-\beta)m_r = 1.2$, greater than 1, so in fact we are under the hypotheses of Conjecture 17.2(iii) and a positive probability of coexistence is predicted.

We have simulated 80 generations of an X-linked two-sex branching process, starting with $Z_0^{RR\times R} = 0$ and $Z_0^{Rr\times R} = 5$ and taking as reproduction laws for $RR \times R$ and $Rr \times R$ couples Poisson distributions with the means m_R and m_r, respectively, as indicated above. In Fig. 17.1 left it is shown that the ratio $Z_n^{RR\times R}/(nZ_n^{Rr\times R})$ approaches to $\beta/(1-\beta) = 2.33$ as n tends to ∞. The middle and right graphs are plots of $\{\log(Z_n^{RR\times R})\}_{n=0}^{\infty}$ and $\{\log(Z_n^{Rr\times R})\}_{n=0}^{\infty}$, respectively. Whereas the first sequence seems to grow parallel to the graph of the function $\log(n) + (n-1)\log((1-\alpha)(1-\beta)m_r)$, the second one follows a line with slope $\log((1-\alpha)(1-\beta)m_r)$, which shows a growth of the number of $RR \times R$ couples with

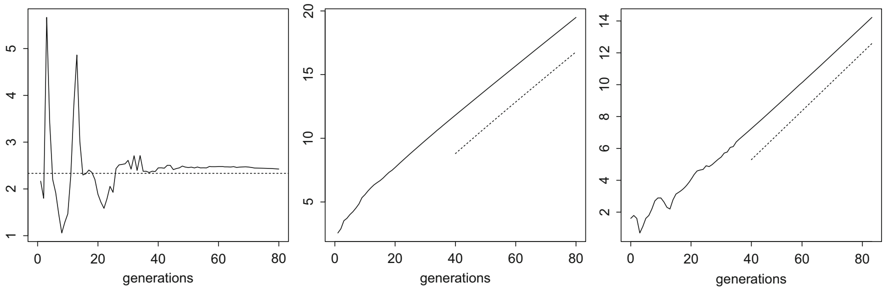


Fig. 17.1 Realizations of $Z_n^{RR\times R}/(nZ_n^{Rr\times R})$ (*left plot*), and $\log(Z_n^{RR\times R})$ (*middle plot*) and $\log(Z_n^{Rr\times R})$ (*right plot*) in a process, where both type of mating units have survived until generation 80

a rate of $n((1-\alpha)(1-\beta)m_r)^{n-1}$, slightly faster than the growth of the number of $Rr \times R$ couples which is determined by $((1-\alpha)(1-\beta)m_r)^n$.

17.6.2 Case $\alpha < 0.5$

If $\alpha < 0.5$, the population balances eventually towards males, but since r males do not participate in the mating, then we have to investigate when the number of R males exceeds that of females in the long term.

Notice first that if the number of R males was eventually greater than the number of females, Conjecture 17.1 seems to be still valid. Indeed, since the number of couples would be determined by the number of females, with a reasoning similar to that of previous subsection, we get for n large enough,

$$Z_{n+1}^{RR\times R} \simeq F_{n+1}^{RR,RR\times R} + F_{n+1}^{RR,Rr\times R} \simeq \alpha m_R Z_n^{RR\times R} + \alpha\beta m_r Z_n^{Rr\times R}, \tag{17.5}$$

$$Z_{n+1}^{Rr\times R} \simeq F_{n+1}^{Rr,Rr\times R} \simeq \alpha(1-\beta)m_r Z_n^{Rr\times R}, \tag{17.6}$$

and the quotient between these expressions leads again to (17.3), from which we derive Conjecture 17.1.

So the key to find out the rates of growth of $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$, and therefore the coexistence of both genotypes is the balance between sexes jointly with Conjecture 17.1. According to this conjecture, if $(1-\beta)m_r \leq m_R$, $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ would be negligible with respect to $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and the production of offspring by $Rr \times R$ couples would not be relevant to establish the balance between sexes. Moreover, since $\alpha < 0.5$ and $RR \times R$ couples only produce R males, the average production of R males per couple exceeds that of females, which eventually would cause a greater R number of males than females.

Therefore, in the case $(1-\beta)m_r < m_R$, (17.5) and (17.6) can be used to derive that the rates of growth of $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ are $(\alpha m_R)^n$ and $(\alpha(1-\beta)m_r)^n$, respectively. In the case $(1-\beta)m_r = m_R$, using the same arguments as for $\alpha \geq 0.5$, the rates of growth that one could expect for $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ are $(\alpha(1-\beta)m_r)^n$ and $n(\alpha(1-\beta)m_r)^{n-1}$, respectively.

A more complex situation seems to be $(1-\beta)m_r > m_R$, because in the long term it is possible both, that the number of R males exceed that of females and that the number of females exceed that of R males. Since $(1-\beta)m_r > m_R$, by Conjecture 17.1, $Z_n^{RR\times R} \simeq \gamma Z_n^{Rr\times R}$ with γ finite and therefore, taking into account (17.4), it seems reasonable to guess that $F_n \leq M_n^R$ if and only if

$$\alpha m_R \gamma + \alpha m_r \leq (1-\alpha)m_R \gamma + (1-\alpha)\beta m_r,$$

or equivalently

$$(2\alpha - 1)m_R \gamma + \alpha m_r \leq (1-\alpha)\beta m_r. \tag{17.7}$$

Notice that, since $\alpha < 0.5$, $2\alpha - 1 < 0$ so (17.7) is a weaker condition than $\alpha m_r \leq (1-\alpha)\beta m_r$ in order to guarantee that R males exceeds females in the long term.

Therefore, if (17.7) holds, we can use expressions (17.5) and (17.6) which, jointly with Conjecture 17.1, lead us to guess that $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ have the same rate of growth, which is $(\alpha(1-\beta)m_r)^n$.

Finally, if (17.7) does not hold, then the number of females exceeds in the long term that of R males and the behaviour of $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ should be analogous to the case $\alpha \geq 0.5$, specifically the same as in Conjecture 17.2(i), with a rate of growth of τ^n for both sequences.

We summarize all these guesses in the following conjecture

Conjecture 17.3 Assume $\alpha < 0.5$ and $\min\{\alpha(1-\beta)m_r, (1-\alpha)(1-\beta)m_r\} > 1$, and let $i \geq 0, j > 0$.

(i) If $(1-\beta)m_r > m_R$, then

  a. If $(2\alpha-1)\gamma m_R+\alpha m_r \leq (1-\alpha)\beta m_r$ then $P_{(i,j)}(Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty) > 0$. Moreover $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ grow at a rate of $(\alpha(1-\beta)m_r)^n$.
  b. If $(2\alpha-1)\gamma m_R + \alpha m_r > (1-\alpha)\beta m_r$, then
  $P_{(i,j)}(Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty) > 0$ if and only if $\tau > 1$.
  Moreover, if $\tau > 1$, $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ grow at a rate of τ^n.

(ii) If $(1-\beta)m_r < m_R$, then $P_{(i,j)}(Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty) > 0$. Moreover, $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ grows at a rate of $(\alpha m_R)^n$ and $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ grows at a rate of $(\alpha(1-\beta)m_r)^n$.

(iii) If $(1-\beta)m_r = m_R$, then $P_{(i,j)}(Z_n^{RR\times R} \to \infty, Z_n^{Rr\times R} \to \infty) > 0$. Moreover, $\{Z_n^{Rr\times R}\}_{n=0}^{\infty}$ grows at a rate of $(\alpha(1-\beta)m_r)^n$ and $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ grows at a rate of $n(\alpha(1-\beta)m_r)^{n-1}$.

Notice that, according to Conjecture 17.3(i)b, if $\tau \leq 1$, $(1-\beta)m_r > m_R$ and $(2\alpha-1)\gamma m_R+\alpha m_r > (1-\alpha)\beta m_r$, then the situation is analogous to that of Remark 17.3. Now, in case of fixation of R allele, the rate of growth of $\{Z_n^{RR\times R}\}_{n=0}^{\infty}$ is $(\alpha m_R)^n$.

Next we show a simulated example that illustrate Conjecture 17.3(i)b. Specifically we have simulated an X-linked two-sex branching process with $\alpha = 0.4$ (biased sex ratios are not unusual, see for example [2]) , $\beta = 0.6$, $m_R = 0.75$ and $m_r = 8$. With these values for the parameters we have that $\alpha(1-\beta)m_r = 1.28$, greater than 1. Also, $0.75 = m_R < (1-\beta)m_r = 3.2$, so the conditions of Conjecture 17.1(i) are satisfied with $\gamma = 1.96$. Moreover $2.906 = (2\alpha - 1)\gamma m_R + \alpha m_r > (1-\alpha)\beta m_r = 2.88$, so in fact we are under the hypotheses of Conjecture 17.3(i)b with $\tau = 1.271$, greater than 1, so a positive probability of coexistence is predicted.

We have simulated 50 generations of an X-linked two-sex branching process, starting with $Z_0^{RR\times R} = 0$ and $Z_0^{Rr\times R} = 5$ and taking as reproduction laws for $RR \times R$ and $Rr \times R$ couples Poisson distributions with the means m_R and m_r, respectively, as indicated above. In Fig. 17.2 left it is shown that the ratio between the number of couples of both types approaches to γ. The middle and right graphs are plots of $\{\log(Z_n^{RR\times R})\}_{n=0}^{\infty}$ and $\{\log(Z_n^{Rr\times R})\}_{n=0}^{\infty}$, respectively. Both sequences seem to grow

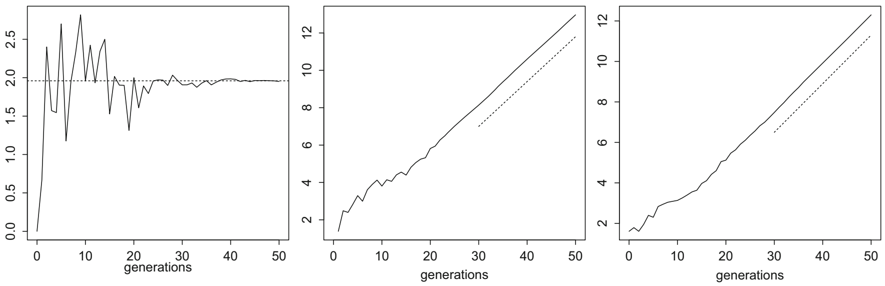


Fig. 17.2 Realizations of $Z_n^{RR\times R}/Z_n^{Rr\times R}$ (*left plot*), and $\log(Z_n^{RR\times R})$ (*middle plot*) and $\log(Z_n^{Rr\times R})$ (*right plot*) in a process, where both type of mating units have survived until generation 50

following a line (dotted line) with slope $\log(\tau)$, which shows the geometric growth in the number of couples of both types, with a common growth rate defined by τ.

Remark 17.4 For the computer simulations, we used the language and environment for statistical computing and graphics **R** ("GNU S") (see [11]).

17.7 Proofs

We first provide a necessary basic notation to prove the results. We start by denoting as $T_{n+1}^{RR\times R}$ and $T_{n+1}^{Rr\times R}$ the total number of individuals in generation $n+1$ generated, respectively, by the $RR\times R$ and $Rr\times R$ couples, $n\geq 0$:

$$T_{n+1}^{RR\times R} = \sum_{i=1}^{Z_n^{RR\times R}} (F_{ni}^{RR,RR\times R} + M_{ni}^{R,RR\times R}) \quad \text{and}$$

$$T_{n+1}^{Rr\times R} = \sum_{i=1}^{Z_n^{Rr\times R}} (F_{ni}^{RR,Rr\times R} + F_{ni}^{Rr,Rr\times R} + M_{ni}^{R,Rr\times R} + M_{ni}^{r,Rr\times R}).$$

We can deduce from the expressions above that the total number of individuals in the $(n+1)$th generation is

$$T_{n+1} = T_{n+1}^{RR\times R} + T_{n+1}^{Rr\times R} = F_{n+1} + M_{n+1},\ n\geq 0.$$

It is easy to prove that the distributions of F_{n+1} and M_{n+1} given T_{n+1} are binomials with parameters (T_{n+1},α) and $(T_{n+1},1-\alpha)$, respectively.

We denote by $f_R(s)$ and $f_r(s)$ the probability generating functions of the $RR\times R$ and $Rr\times R$ couple reproduction laws, respectively. Recall that those reproduction laws have finite means and variances.

Finally, we introduce the σ-algebras $\mathcal{G}_0 = \sigma(Z_0^{RR\times R}, Z_0^{Rr\times R})$,

$$\mathcal{G}_n = \sigma(Z_0^{RR\times R}, Z_0^{Rr\times R}, F_k^{RR}, F_k^{Rr}, M_k^R, M_k^r, Z_k^{RR\times R}, Z_k^{Rr\times R}, k = 1, \ldots, n),\ n \geq 1,$$

and $\mathcal{F}_n = \sigma(\mathcal{G}_{n-1}, F_n^{RR}, F_n^{Rr}, M_n^R, M_n^r),\ n \geq 1$. For any $i, j \geq 0$, recall that we write $P_{(i,j)}(\cdot)$ for $P(\cdot|Z_0^{RR\times R} = i, Z_0^{Rr\times R} = j)$, and now we introduce the notation $E_{(i,j)}[\cdot] = E[\cdot|Z_0^{RR\times R} = i, Z_0^{Rr\times R} = j]$.

17.7.1 Proof of Theorem 17.1

In order to prove that $P(Z_n^{Rr\times R} \to 0) + P(Z_n^{Rr\times R} \to \infty) = 1$, we prove that the probability of the complementary set is equal to 0. For that, it is enough to prove that for all $i' \geq 1$ and $i, j \geq 0$

$$P_{(i,j)}(Z_n^{Rr\times R} = i',\ \text{i.o.}) = 0,$$

where i.o. stands for infinitely often. Taking $a = P(F_{01}^{Rr,Rr\times R} = 0)$, since $0 < \alpha, \beta < 1$, and taking into account the multinomial scheme assumed in the reproduction of the $Rr \times R$ couples, then $a = f_r(\alpha\beta + (1-\alpha))$ is positive and constant. We conclude analogously to the proof of Theorem 3.1 in [5] that

$$\begin{aligned} P_{(i,j)}(Z_n^{Rr\times R} = i',\ \text{i.o.}) &= \lim_{m\to\infty} P_{(i,j)}(Z_n^{Rr\times R} = i' \text{ for at least } m \text{ values of } n > 0) \\ &\leq \lim_{m\to\infty} (1 - a^i)(1 - a^{i'})^{m-1} = 0. \end{aligned}$$

17.7.2 Proof of Theorem 17.2

By Theorem 17.1, it is enough to prove that $P_{(i,j)}(Z_n^{Rr\times R} \to \infty) = 0$.

(i) Assume $\alpha(1-\beta)m_r \leq 1$. From the definition of the model, $Z_n^{Rr\times R} \leq F_n^{Rr}$ for all $n \geq 1$, and consequently, by using Proposition 17.1(iii) we obtain

$$E[Z_n^{Rr\times R}|\mathcal{G}_{n-1}] \leq E[F_n^{Rr}|\mathcal{G}_{n-1}] = \alpha(1-\beta)m_r Z_{n-1}^{Rr\times R} \leq Z_{n-1}^{Rr\times R} \text{ a.s.},$$

and one concludes that $\{Z_n^{Rr\times R}\}_{n\geq 0}$ is a non-negative supermartingale, and so converges to a finite limit. Hence $P_{(i,j)}(Z_n^{Rr\times R} \to \infty) = 0$.

(ii) Assume $0 < (1-\alpha)(1-\beta)m_r < 1$ and $\alpha > 0.5$ [otherwise the result is deduced from (i)].

For each $N > 0$ and for some positive constant A, we define the sequences of sets $B_k = \{Z_{N+k}^{Rr\times R} \leq A\}$, $k \geq 0$, and the stopping time

$$T_N = \begin{cases} \infty & \text{if } \inf_{k\geq N} Z_k^{Rr\times R} > A \\ \min\{k \geq N : Z_k^{Rr\times R} \leq A\} & \text{otherwise} \end{cases}$$

such that $B_k \subseteq \{T_N \leq N + k\}$.

Consider also the sequence of σ-algebras $\{\mathcal{G}_n\}_{n\geq 0}$ as defined above. If we prove that, for all $k \geq 1$,

$$E[Z_{N+k}^{Rr\times R}|\mathcal{G}_{N+k-1}] \leq Z_{N+k-1}^{Rr\times R} \text{ a.s. on } \{Z_{N+k-1}^{Rr\times R} > A\}, \tag{17.8}$$

applying the Lemma 3 in [6], we obtain that $\{Z_{T_N\wedge(N+k)}^{Rr\times R}\}_{k\geq 0}$ is a non-negative supermartingale, with $T_N \wedge (N+k) = \min\{T_N, N+k\}$, and consequently converges to a non-negative and finite limit. But, for every N, on the set $\{\inf\limits_{k\geq N} Z_k^{Rr\times R} > A\}$, $Z_{T_N\wedge(N+k)}^{Rr\times R} = Z_{N+k}^{Rr\times R}$ a.s. for all $k \geq 0$, so the sequence on right-hand side also converges to a non-negative and finite limit, hence

$$P_{(i,j)}(\{Z_{N+k}^{Rr\times R} \to \infty\} \cap \{\inf_{k\geq N} Z_k^{Rr\times R} > A\}) = 0,$$

and this would conclude the proof.

In order to prove (17.8), we simplify the notation and write $n = N + k$. Taking into account that $\mathcal{G}_{n-1} \subseteq \mathcal{F}_n, n \geq 1$ and applying Proposition 17.1 (i), we have a.s.

$$\begin{aligned} E[Z_n^{Rr\times R}|\mathcal{G}_{n-1}] &= E[E[Z_n^{Rr\times R}|\mathcal{F}_n]|\mathcal{G}_{n-1}] \\ &= E[F_n^{Rr} I_{\{F_n\leq M_n^R\}}|\mathcal{G}_{n-1}] + E\left[\frac{M_n^R}{F_n} F_n^{Rr} I_{\{F_n>M_n^R\}}|\mathcal{G}_{n-1}\right]. \end{aligned} \tag{17.9}$$

Let us bound properly each of these summands.

For the first summand of (17.9), since we are assuming that M_n given T_n is distributed according to a binomial scheme, we can apply a Chernoff type of inequality, and have, for all n and $l > 0$,

$$P(F_n \leq M_n|T_n = l) = P(M_n \geq l/2|T_n = l) \leq (2\sqrt{\alpha(1-\alpha)})^l = a^l,$$

with $a = 2\sqrt{\alpha(1-\alpha)} < 1$ as $\alpha > 0.5$. Then, due to the mutual independence of the $RR \times R$ and $Rr \times R$ reproduction laws and that $M_n^R \leq M_n$, $n \geq 1$,

$$\begin{aligned} P(F_n \leq M_n^R|\mathcal{G}_{n-1}) &\leq P(F_n \leq M_n|\mathcal{G}_{n-1}) \\ &= E[P(F_n \leq M_n|T_n)|\mathcal{G}_{n-1}] \\ &\leq E[a^{T_n}|\mathcal{G}_{n-1}] = E[a^{T_n^{RR\times R}}|\mathcal{G}_{n-1}]E[a^{T_n^{Rr\times R}}|\mathcal{G}_{n-1}] \\ &= f_R(a)^{Z_{n-1}^{RR\times R}} f_r(a)^{Z_{n-1}^{Rr\times R}} \leq f_r(a)^{Z_{n-1}^{Rr\times R}} \text{ a.s.} \end{aligned} \tag{17.10}$$

Therefore, applying (17.10), the Cauchy-Schwartz inequality, and taking into account that the variances are assumed finite,

$$\begin{aligned} E[F_n^{Rr} I_{\{F_n \le M_n^R\}} | \mathcal{G}_{n-1}] &\le E[(F_n^{Rr})^2 | \mathcal{G}_{n-1}]^{1/2} P(F_n \le M_n^R | \mathcal{G}_{n-1})^{1/2} \\ &\le Z_{n-1}^{Rr\times R} f_r(a)^{Z_{n-1}^{Rr\times R}/2} \quad \text{a.s.,} \end{aligned} \tag{17.11}$$

for some positive constant K_1.

To bound the second summand of (17.9), given $\varepsilon > 0$, define $\gamma_1 = (1-\alpha)(1-\beta)(m_r + \varepsilon)$, and $B_\varepsilon = 1 + 2\varepsilon/m$, where $m = \min\{m_R, m_r\} - \varepsilon$ if $m_R \neq 0$ or $m = m_r - \varepsilon$ otherwise. We take ε small enough such that $0 < \gamma_1 B_\varepsilon < 1$, $m > 0$, $m_R - \varepsilon > 0$ (if $m_R \neq 0$) and $\beta m_r - \varepsilon > 0$.

For each $n \ge 1$, define the sets

$$\begin{aligned} A_{F,n} &= \{|F_n - (\alpha m_R Z_{n-1}^{RR\times R} + \alpha m_r Z_{n-1}^{Rr\times R})| \le \alpha\varepsilon Z_{n-1}\}, \\ A_{M^R,n} &= \{|M_n^R - ((1-\alpha) m_R Z_{n-1}^{RR\times R} + (1-\alpha)\beta m_r Z_{n-1}^{Rr\times R})| \le (1-\alpha)\varepsilon Z_{n-1}\}, \\ A_{F^{Rr},n} &= \{|F_n^{Rr} - \alpha(1-\beta) m_r Z_{n-1}^{Rr\times R}| \le \alpha(1-\beta)\varepsilon Z_{n-1}^{Rr\times R}\}. \end{aligned}$$

Since there is no ambiguity, from now on n is dropped from the notation of these sets. Define also $D = A_F \cap A_{M^R} \cap A_{F^{Rr}}$. With this notation we write

$$\begin{aligned} & E\left[\frac{M_n^R}{F_n} F_n^{Rr} I_{\{F_n > M_n^R\}} | \mathcal{G}_{n-1}\right] \\ &= E\left[\frac{M_n^R}{F_n} F_n^{Rr} I_{\{F_n > M_n^R\}} I_{D^c} | \mathcal{G}_{n-1}\right] + E\left[\frac{M_n^R}{F_n} F_n^{Rr} I_{\{F_n > M_n^R\}} I_D | \mathcal{G}_{n-1}\right] \quad \text{a.s.} \end{aligned} \tag{17.12}$$

Since the reproduction laws are assumed to have finite variances, an immediate application of Proposition 17.1 and Chebyshev's inequality gives

$$P(A_F^c | \mathcal{G}_{n-1}) \le \frac{C_1}{Z_{n-1}} \le \frac{C_1}{Z_{n-1}^{Rr\times R}} \text{ a.s.,} \qquad P(A_{M^R}^c | \mathcal{G}_{n-1}) \le \frac{C_2}{Z_{n-1}} \le \frac{C_2}{Z_{n-1}^{Rr\times R}} \text{ a.s.}$$

and

$$P(A_{F^{Rr}}^c | \mathcal{G}_{n-1}) \le \frac{C_3}{Z_{n-1}^{Rr\times R}} \text{ a.s.,}$$

for certain positive constants C_1, C_2, and C_3. Therefore, for some positive constant C_4,

$$P(D^c | \mathcal{G}_{n-1}) \le \frac{C_4}{Z_{n-1}^{Rr\times R}} \text{ a.s.} \tag{17.13}$$

Now, applying (17.13) and the Cauchy-Schwartz inequality,

$$\begin{aligned} E\left[\frac{M_n^R}{F_n}F_n^{Rr}I_{\{F_n>M_n^R\}}I_{D^c}|\mathcal{G}_{n-1}\right] &\leq E[F_n^{Rr}I_{D^c}|\mathcal{G}_{n-1}] \\ &\leq E[(F_n^{Rr})^2|\mathcal{G}_{n-1}]^{1/2}P(D^c|\mathcal{G}_{n-1})^{1/2} \\ &\leq K_2 Z_{n-1}^{Rr\times R}(C_5 Z_{n-1}^{Rr\times R})^{-1/2} = K_3(Z_{n-1}^{Rr\times R})^{1/2} \quad \text{a.s.,} \end{aligned} \tag{17.14}$$

for some positive constants K_2, C_5, and K_3. Finally, on D, if $m_R \neq 0$,

$$\begin{aligned} & E\left[\frac{M_n^R}{F_n}F_n^{Rr}I_{\{F_n>M_n^R\}}I_D|\mathcal{G}_{n-1}\right] \\ \leq\ & \alpha(1-\beta)(m_r+\varepsilon)Z_{n-1}^{Rr\times R}\frac{(1-\alpha)((m_R+\varepsilon)Z_{n-1}^{RR\times R}+(\beta m_r+\varepsilon)Z_{n-1}^{Rr\times R})}{\alpha((m_R-\varepsilon)Z_{n-1}^{RR\times R}+(m_r-\varepsilon)Z_{n-1}^{Rr\times R})} \\ \leq\ & (1-\alpha)(1-\beta)(m_r+\varepsilon)Z_{n-1}^{Rr\times R}\frac{(m_R-\varepsilon)Z_{n-1}^{RR\times R}+2\varepsilon Z_{n-1}^{RR\times R}+(m_r-\varepsilon)Z_{n-1}^{Rr\times R}+2\varepsilon Z_{n-1}^{Rr\times R}}{(m_R-\varepsilon)Z_{n-1}^{RR\times R}+(m_r-\varepsilon)Z_{n-1}^{Rr\times R}} \\ \leq\ & \gamma_1 Z_{n-1}^{Rr\times R}\left(1+\frac{2\varepsilon}{m}\right) = \gamma_1 B_\varepsilon Z_{n-1}^{Rr\times R} \quad \text{a.s.,} \end{aligned} \tag{17.15}$$

with $m = \min\{m_R, m_r\} - \varepsilon$. And, if $m_R = 0$,

$$\begin{aligned} E\left[\frac{M_n^R}{F_n}F_n^{Rr}I_{\{F_n>M_n^R\}}I_D|\mathcal{G}_{n-1}\right] &\leq \alpha(1-\beta)(m_r+\varepsilon)Z_{n-1}^{Rr\times R}\frac{(1-\alpha)(\beta m_r+\varepsilon)Z_{n-1}^{Rr\times R}}{\alpha(m_r-\varepsilon)Z_{n-1}^{Rr\times R}} \\ &\leq \gamma_1 Z_{n-1}^{Rr\times R}\left(1+\frac{2\varepsilon}{m}\right) = \gamma_1 B_\varepsilon Z_{n-1}^{Rr\times R} \quad \text{a.s.,} \end{aligned} \tag{17.16}$$

with $m = m_r - \varepsilon$. In any case,

$$E\left[\frac{M_n^R}{F_n}F_n^{Rr}I_{\{F_n>M_n^R\}}I_D|\mathcal{G}_{n-1}\right] \leq \gamma_1 B_\varepsilon Z_{n-1}^{Rr\times R} \quad \text{a.s.} \tag{17.17}$$

Summarizing, from (17.11), (17.12), (17.15) and (17.17), we deduce that

$$E[Z_n^{Rr\times R}|\mathcal{G}_{n-1}] \leq (K_1 f_r(a)^{Z_{n-1}^{Rr\times R}/2} + K_3(Z_{n-1}^{Rr\times R})^{-1/2} + \gamma_1 B_\varepsilon)Z_{n-1}^{Rr\times R} \quad \text{a.s.}$$

Recall we have taken ε such that $\gamma_1 B_\varepsilon < 1$. Hence, since $f_r(a) < 1$, we can take $A > 0$ large enough such that, for $Z_{n-1}^{Rr\times R} > A$,

$$K_1 f_r(a)^{Z_{n-1}^{Rr\times R}/2} < 1 - \gamma_1 B_\varepsilon$$

and consequently the term in parentheses is less than 1. For this value of A, (17.8) holds, and therefore the result is proved.

Acknowledgements This research has been supported by the Ministerio de Economía y Competitividad of Spain (grants MTM2012-31235 and MTM2015-70522-P), the Junta de Extremadura (grant GR15105), the FEDER, and the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria (grant DFNI-I02/17).

References

1. Alsmeyer, G., Gutiérrez, C., Martínez, R.: Limiting genotype frequencies of Y-linked genes through bisexual branching processes with blind choice. J. Theor. Biol. **275**, 42–51 (2011)
2. Bonal, R., Hernández, M., Espelta, J.M., Muñoz, A., Aparicio, J.M.: Unexpected consequences of a drier world: evidence that delay in late summer rains biases the population sex ratio of an insect. R. Soc. Open Sci. **2**, 150198 (2015)
3. Daley, D.J.: Extinction conditions for certain bisexual Galton-Watson branching processes. Z. Wahrscheinlichkeitsth. **9**, 315–322 (1968)
4. Germain, D. P.: General aspects of X-linked diseases. In: Mehta, A., Beck, M., Sunder-Plassmann, G. (eds.) Fabry Disease Perspectives from 5 Years of FOS, Chap. 7. Oxford PharmaGenesis, Oxford (2006). http://www.ncbi.nlm.nih.gov/books/NBK11593/
5. González, M., Hull, D.M., Martínez, R., Mota, M.: Bisexual branching processes in a genetic context: the extinction problem for Y-linked genes. Math. Biosci. **202**, 227–247 (2006)
6. González, M., Martínez, R., Mota, M.: Bisexual branching processes in a genetic context: rates of growth for Y-linked genes. Math. Biosci. **215**, 167–176 (2008)
7. González, M., Gutiérrez, C., Martínez, R.: Parametric inference for Y-linked gene branching models: expectation-maximization method. In: Workshop on Branching Processes and Their Applications. Lecture Notes in Statistics, vol. 197, pp. 191–204. Springer, Heidelberg (2010)
8. González, M., Gutiérrez, C., Martínez, R.: Expectation-maximization algorithm for determining natural selection of Y-linked genes through two-sex branching processes. J. Comput. Biol. **19**, 1015–1026 (2012)
9. González, M., Gutiérrez, C., Martínez, R.: Extinction conditions for Y-linked mutant-alleles through two-sex branching processes with blind-mating structure. J. Theor. Biol. **307**, 104–116 (2012)
10. González, M., Gutiérrez, C., Martínez, R.: Parametric Bayesian inference for Y-linked two-sex branching models. Stat. Comput. **23**, 727–741 (2013)
11. R Development Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/ (2015). ISBN 3-900051-07-0

Chapter 18
Two-Sex Branching Processes with Several Mating and Reproduction Strategies: Extinction Versus Survival

Manuel Molina, Manuel Mota, and Alfonso Ramos

Mathematics Subject Classification (2000): 60J80

18.1 Introduction

In the general context of stochastic modeling, the branching process theory provides appropriate mathematical models to describe the probabilistic evolution of biological populations whose sizes vary over time, due to random births and deaths, see for details the monographs [3] or [5].

We focus here our interest on the development of stochastic processes to describe the demographic dynamics of biological populations with sexual reproduction. This research line was initiated by Daley [1] who introduced the bisexual Galton-Watson branching process. By considering such a process as mathematical model and assuming some specific mating strategies, he established conditions for the extinction/survival of the population.

Taking Daley's process as starting point, some population-size dependent two-sex branching processes have been investigated in the literature [6–8] or [10]. These processes describe populations where several mating and reproduction strategies are feasible and, in each generation, one of them is implemented. The choice of the mating and/or reproduction strategy is assumed to depend on the number of couples

---

M. Molina (✉) • M. Mota
Faculty of Sciences, Department of Mathematics, University of Extremadura, Avda. Elvas, 06006 Badajoz, Spain
e-mail: mmolina@unex.es; mota@unex.es

A. Ramos
Facultad de Veterinaria, Department of Mathematics, University of Extremadura, Cáceres, Spain
e-mail: aramos@unex.es

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_18

female-male in the population. However, it seems more realistic that mating and reproduction (or both) are affected by the numbers of females and males in the population. This novelty has been considered in a recent work by Molina et al. [9] where a female-male population-size dependent two-sex branching process has been introduced and, from a Bayesian outlook, some inferential problems have been investigated.

The motivation behind this work is to continue the research about that new two-sex branching process from a more theoretical point of view. Specifically, the main results of the paper are related to the possible extinction/survival of populations modeled by such a process. Previously, we provide a detailed description of the probabilistic model.

18.2 The Model

Let us consider populations with sexual reproduction, i.e. individuals must form couples female-male before reproduction. Several strategies for the mating are feasible, depending on the current numbers of females and males. One of them is implemented in each generation. Once formed the couples, there are several reproduction strategies to produce new descendants. The choice of the reproduction strategy depends also on the current numbers of females and males. Assume also that in this reproduction phase the couples act independently of the others and, inside a generation, all of them adopt the same strategy.

Denote by N_m and N_r the positive integers representing the numbers of mating and reproduction strategies, respectively. We model mathematically the demographic dynamics of such biological populations as follows:

1. Let $L_l, l = 1, \dots, N_m$, be functions, defined on $\mathbb{N}^2$ and taking values in $\mathbb{N}$, where $\mathbb{N}$ denotes the set of non-negative integers. The role of L_l is to determine the number of couples female-male formed in the population when the lth mating strategy is considered. Each function L_l is assumed to be non-decreasing and such that, $L_l(f, 0) = L_l(0, m) = 0, f, m \in \mathbb{N}$ (if there are either no females or no males in the population then it is not possible to form couples). These functions will be referred as to mating functions.
2. Let $\{p^h_{k,j}\}_{(k,j)\in\mathbb{N}^2}$, $h = 1, \dots, N_r$, be offspring probability distributions, $p^h_{k,j}$ representing the probability that a couple produces k females and j males when the hth reproduction strategy is considered. We denote by $\boldsymbol{\mu}^h$ and Σ^h, respectively, the mean vector and the covariance matrix of the hth offspring probability distribution.
3. Let ψ and φ be functions defined on $\mathbb{N}^2$ and taking values in the sets $\{1, \dots, N_m\}$ and $\{1, \dots, N_r\}$ respectively. Their role is to determine, according to the numbers of females and males in the population, the mating function and the offspring probability distribution to be considered in the mating and reproduction phases, respectively. In fact, if there are f females and m males in the population then

L_l and $\{p^h_{k,j}\}_{(k,j)\in\mathbb{N}^2}$, where $l = \psi(f,m)$ and $h = \varphi(f,m)$, are the corresponding mating function and offspring probability distribution to be considered.

4. Let us define the two-sex branching process $\{\mathbf{X}_n\}_{n=0}^{\infty}$, where $\mathbf{X}_n = (F_n, M_n)$ is the random vector representing the numbers of females (F_n) and males (M_n) at time (generation) n, in the form:

Assume that in a given generation n, the total number of females and males in the population is given by $\mathbf{X}_n = (F_n, M_n)$. Then, the mating strategy considered is the strategy number $l_n := \psi(\mathbf{X}_n)$. Consequently, $Z_n := L_{l_n}(\mathbf{X}_n)$ couples are formed. All these couples adopt the reproduction strategy number $h_n := \varphi(\mathbf{X}_n)$. Each of them produces, independently of the others, a random number of females and males, according to the offspring probability distribution $\{p^{h_n}_{k,j}\}_{(k,j)\in\mathbb{N}^2}$ and then disappears. As consequence, the vector representing the numbers of females and males in the $(n+1)$st generation, namely $\mathbf{X}_{n+1} = (F_{n+1}, M_{n+1})$, is determined as

$$\mathbf{X}_{n+1} := \sum_{i=1}^{Z_n} \left(f^{h_n}_{n,i}, m^{h_n}_{n,i} \right),$$

where $(f^{h_n}_{n,i}, m^{h_n}_{n,i})$ are independent and identically distributed random variables such that

$$P(f^{h_n}_{n,1} = k, m^{h_n}_{n,1} = j) = p^{h_n}_{k,j}, \ (k,j) \in \mathbb{N}^2.$$

Initially, we assume f_0 females and m_0 males in the population, i.e. $\mathbf{X}_0 = (f_0, m_0) \in \mathbb{N}^2_+$, where $\mathbb{N}_+$ denotes the positive integers.

It can be easily verified that $\{\mathbf{X}_n\}_{n=0}^{\infty}$ is a homogeneous Markov chain with state space on $\mathbb{N}^2$. In fact, if $\mathbf{x}_n, \mathbf{x}_{n+1} \in \mathbb{N}^2$, $l = \psi(\mathbf{x}_n)$ and $h = \varphi(\mathbf{x}_n)$, then

$$P(\mathbf{X}_{n+1} = \mathbf{x}_{n+1} | \mathbf{X}_n = \mathbf{x}_n) = P\left(\sum_{i=1}^{L_l(\mathbf{x}_n)} (f^h_{n,i}, m^h_{n,i}) = \mathbf{x}_{n+1} \right) = \left(P^h_{\mathbf{x}_{n+1}} \right)^{*L_l(\mathbf{x}_n)}$$

with $(P^h_{\mathbf{x}})^{*j}$, $j \in \mathbb{N}$, $\mathbf{x} \in \mathbb{N}^2$, representing the conventional convolution notation concerning the offspring probability distribution $\{p^h_{k,j}\}_{(k,j)\in\mathbb{N}^2}$.

Remark 18.1 In addition to its theoretical interest, the two-sex branching process described above has also practical implications. Really, it is appropriate to model the probabilistic evolution of animal species characterized by a single reproductive episode before death (semelparous species). For the particular case when only one strategy is considered in the respective mating and reproduction phases, namely $N_r = N_m = 1$, it is reduced to the bisexual Galton-Watson branching process introduced in [1].

In the sequel, we shall study several theoretical questions about the two-sex branching process described in Sect. 18.2. Specifically, Sect. 18.3 is devoted to establish some basic properties about the underlying probabilistic model. Then, in Sect. 18.4, we include the main results of the work concerning the possible extinction/survival of the population. To this end, we deal with the sequence $\{g(\mathbf{X}_n)\}_{n=0}^{\infty}$ where $g(\cdot)$ has the functional form

$$g(f,m) = a_1 f + a_2 m,\ \ a_1 \geq 0,\ \ a_2 \geq 0,\ \ a_1 + a_2 > 0,\ \ (f,m) \in \mathbb{N}^2. \tag{18.1}$$

Some particular cases of biological interest are $g(f,m) = f$, $g(f,m) = m$, or $g(f,m) = f + m$, corresponding to the numbers of females, males, or total individuals, respectively.

18.3 Basic Properties

In this section we provide three results about the two-sex branching process introduced. First, we determine the first and second order conditional moments of the process. Then, in a second result, we establish its Markovian properties. Finally, we use this last result to derive some conditions which guarantee the classical duality extinction-explosion property in branching process theory.

Proposition 18.1 *Let us consider the process* $\{\mathbf{X}_n\}_{n=0}^{\infty}$*. Let* $\mathbf{x} \in \mathbb{N}^2$*,* $l = \psi(\mathbf{x})$*,* $h = \varphi(\mathbf{x})$*, and* $g(\cdot)$ *as in (18.1). Then*

(i) $E[\mathbf{X_{n+1}}|\mathbf{X_n} = \mathbf{x}] = L_l(\mathbf{x})\mu^h$.
(ii) $Var[\mathbf{X_{n+1}}|\mathbf{X_n} = \mathbf{x}] = L_l(\mathbf{x})\Sigma^h$.
(iii) $E[g(\mathbf{X_{n+1}})|\mathbf{X_n} = \mathbf{x}] = L_l(\mathbf{x})g(\mu^h)$.
(iv) $Var[g(\mathbf{X_{n+1}})|\mathbf{X_n} = \mathbf{x}] = L_l(\mathbf{x})\mathbf{a}\Sigma^h\mathbf{a}'$*, with* $\mathbf{a} = (a_1, a_2)$ *and* $\mathbf{a}'$ *being the transpose vector of* $\mathbf{a}$.

Proof It is derived from some straightforward calculations.

Definition 18.1 For $\mathbf{x} = (f,m) \in \mathbb{N}_+^2$, $l = \psi(\mathbf{x})$, $h = \varphi(\mathbf{x})$, and $g(\cdot)$ as in (18.1), we define the rate

$$m_g(\mathbf{x}) := \frac{1}{g(\mathbf{x})} E[g(\mathbf{X_{n+1}})|\mathbf{X_n} = \mathbf{x}] = \frac{L_l(\mathbf{x})g(\mu^h)}{g(\mathbf{x})}.$$

Remark 18.2 The rate $m_g(\mathbf{x})$ will play an important role in the study concerning the extinction/survival of the process. Note that, it is the product of two components. The first one ($L_l(\mathbf{x})/g(\mathbf{x})$) is the ratio of couples per unity of "$g(\mathbf{x})$", for instance, ratio of couples per female (when $g(\mathbf{x}) = f$), per male (when $g(\mathbf{x}) = m$), or per individual (when $g(\mathbf{x}) = f + m$). The second one ($g(\mu^h)$) is the expected offspring, in terms of the functional $g(\mathbf{x})$, per couple. For example, expected female offspring (when $g(\mathbf{x}) = f$), expected male offspring (when $g(\mathbf{x}) = m$), or expected total offspring (when $g(\mathbf{x}) = f + m$).

The next result provides a classification of the states of the process. We omit the states $\mathbf{x} \in \mathbb{N}_+^2$ such that $L_l(\mathbf{x}) = 0$, $l = \psi(\mathbf{x})$, which are trivially transient.

Proposition 18.2 *Let us consider the process* $\{\mathbf{X}_n\}_{n=0}^{\infty}$.

(i) $\mathbf{0} = (0,0)$ *is an absorbing state and* $(f,0)$, $(0,m)$, $f,m \in \mathbb{N}_+$ *are transient states.*

(ii) Let $\mathbf{x} \in \mathbb{N}_+^2$ *and* $h = \varphi(\mathbf{x})$. *If*

$$\max\{\sum_{k=0}^{\infty} p_{k,0}^h, \sum_{j=0}^{\infty} p_{0,j}^h\} > 0, \tag{18.2}$$

then $\mathbf{x}$ *is a transient state.*

Proof

(i) If, for some $n \in \mathbb{N}_+$, $\mathbf{X}_n \in \{(f,0),(0,m), f,m \in \mathbb{N}\}$ then $\mathbf{X}_{n+k} = \mathbf{0}$ for all $k \geq 1$. Hence, we deduce that $\mathbf{0}$ is an absorbing state and $(f,0),(0,m), f,m \in \mathbb{N}_+$ are transient states.

(ii) Let $\mathbf{x} \in \mathbb{N}_+^2$, $l = \psi(\mathbf{x})$ and $h = \varphi(\mathbf{x})$.
If (18.2) holds, there exist $k \in \mathbb{N}$ and/or $j \in \mathbb{N}$ such that $\max\{p_{k,0}^h, p_{0,j}^h\} > 0$. Assume that $p_{k,0}^h > 0$. It is sufficient to verify that

$$P(\mathbf{X}_{n+1} = (L_l(\mathbf{x})\, k, 0) | \mathbf{X}_n = \mathbf{x}) > 0.$$

Now

$$P(\mathbf{X}_{n+1} = (L_l(\mathbf{x})\, k, 0) | \mathbf{X}_n = \mathbf{x}) = P\left(\sum_{i=1}^{L_l(\mathbf{x})} (f_{n,i}^h, m_{n,i}^h) = (L_l(\mathbf{x})\, k, 0)\right)$$

$$\geq P((f_{n,i}^h, m_{n,i}^h) = (k,0)\, , i = 1, \ldots, L_l(\mathbf{x})) = (p_{k,0}^h)^{L_l(\mathbf{x})} > 0$$

and the proof is completed. By using a similar reasoning, we deduce the result when it is assumed that $p_{0,j}^h > 0$.

Previous result provides the key to establish the extinction-explosion property for the process $\{\mathbf{X}_n\}_{n=0}^{\infty}$.

Proposition 18.3 *Let us consider the process* $\{\mathbf{X}_n\}_{n=0}^{\infty}$ *and* $g(\cdot)$ *as in (18.1) with* $a_1, a_2 > 0$. *Assume that for all* $h \in \{1, \ldots, N_r\}$

$$\max\{\sum_{k=0}^{\infty} p_{k,0}^h, \sum_{j=0}^{\infty} p_{0,j}^h\} > 0.$$

Then

$$P(\lim_{n\to\infty} \mathbf{X}_n = \mathbf{0}) + P(\lim_{n\to\infty} g(\mathbf{X}_n) = \infty) = 1. \tag{18.3}$$

Proof It is sufficient to prove that, for all $C > 0$,

$$P(\limsup_{n\to\infty}\{0 < g(\mathbf{X}_n) \leq C\}) = 0.$$

Now, writing $\mathbf{x} = (f, m)$, we have that $g(\mathbf{x}) = a_1 f + a_2 m$, with $a_1, a_2 > 0$. Hence, there exists a finite number of states $\mathbf{x} \in \mathbb{N}_+^2$ such that $0 < g(\mathbf{x}) \leq C$. Therefore,

$$P(\limsup_{n\to\infty}\{0 < g(\mathbf{X}_n) \leq C\}) = \sum_{\{\mathbf{x}:\ 0<g(\mathbf{x})\leq C\}} P(\limsup_{n\to\infty}\{\mathbf{X}_n = \mathbf{x}\}). \tag{18.4}$$

By applying Proposition 18.2(b), we derive that all the states $\mathbf{x} \neq \mathbf{0}$ are transient. This implies that

$$P(\limsup_{n\to\infty}\{\mathbf{X}_n = \mathbf{x}\}) = 0.$$

Taking into account (18.4), we conclude the proof.

18.4 Extinction and Survival

We now investigate conditions which guarantee the almost sure extinction of the population and conditions which ensure a positive probability for its survival. First, we introduce the concept of probability of extinction.

Definition 18.2 Let $\mathbf{x}_0 \in \mathbb{N}_+^2$, $\mathbf{x}_0 = (f_0, m_0)$. We define the probability of extinction of the process $\{\mathbf{X}_n\}_{n=0}^{\infty}$, started with f_0 females and m_0 males, by

$$q(\mathbf{x}_0) := P(\lim_{n\to\infty} \mathbf{X}_n = \mathbf{0} | \mathbf{X}_0 = \mathbf{x}_0).$$

Since $\mathbf{0}$ is an absorbing state for $\{\mathbf{X}_n\}_{n=0}^{\infty}$, we derive that

$$q(\mathbf{x}_0) = \lim_{n\to\infty} P(\mathbf{X}_n = \mathbf{0} | \mathbf{X}_0 = \mathbf{x}_0).$$

By using the rate $m_g(\mathbf{x})$ defined in the previous section, we establish the following results.

Theorem 18.1 *Let us consider the process* $\{\mathbf{X}_n\}_{n=0}^{\infty}$ *and* $g(\cdot)$ *as in (18.1) such that (18.3) holds and, for some* $g^* \geq 0$,

$$\sup_{\{\mathbf{x}:\ g(\mathbf{x})>g^*\}} m_g(\mathbf{x}) \leq 1. \tag{18.5}$$

Then $q(\mathbf{x}_0) = 1$ *for all* $\mathbf{x_0} \in \mathbb{N}_+^2$.

Proof Consider the sequence of σ-algebras $\{\mathcal{F}_n\}_{n=0}^{\infty}$, $\mathcal{F}_n := \sigma(\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_n)$. We will develop the proof in two steps.

First step. Assume $g^* = 0$.

Using (18.5) and taking into account Proposition 18.1(b) and Remark 18.2, we deduce that

$$E[g(\mathbf{X}_{\mathbf{n+1}})|\mathcal{F}_n] = m_g(\mathbf{X}_n)g(\mathbf{X}_n) \le g(\mathbf{X}_n). \tag{18.6}$$

Thus $\{g(\mathbf{X}_n)\}_{n=0}^{\infty}$ is a non-negative supermartingale with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$. Hence, by the martingale convergence theorem, there exists a non-negative and finite random variable W such that $\lim_{n\to\infty} g(\mathbf{X}_n) = W$ almost surely. Therefore, for all $\mathbf{x}_0 \in \mathbb{N}_+^2$,

$$P(\limsup_{n\to\infty} g(\mathbf{X}_n) = \infty|\mathbf{X}_0 = \mathbf{x}_0) = 0.$$

From (18.3), we derive that $q(\mathbf{x}_0) = 1$.

Second step. Assume $g^* > 0$.

From the first equality of (18.6), we deduce that $\{g(\mathbf{X}_{n+1})\big(\prod_{k=0}^{n} m_g(\mathbf{X}_k)\big)^{-1}\}_{n=0}^{\infty}$ is a non-negative martingale with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$. Therefore, it converges almost surely to some non-negative random variable W implying that the following equality holds almost surely,

$$g(\mathbf{X}_{n+1}) = W\prod_{k=0}^{n} m_g(\mathbf{X}_k) + R_{n+1}\prod_{k=0}^{n} m_g(\mathbf{X}_k), \tag{18.7}$$

where $\lim\limits_{n\to\infty} R_n = 0$ almost surely. Now, using again (18.3), it is sufficient to prove that

$$P(\lim_{n\to\infty} g(\mathbf{X}_n) = \infty) = 0.$$

Note that, for each trajectory $\omega \in \{\lim\limits_{n\to\infty} g(\mathbf{X}_n) = \infty\}$, there exists n_ω such that $g(\mathbf{X}_k(\omega)) \ge g^*$, for all $k \ge n_\omega$. Consequently, for $n > n_\omega$,

$$\begin{aligned}\prod_{k=0}^{n} m_g(\mathbf{X}_k(\omega)) &\le \prod_{k=0}^{n_\omega} m_g(\mathbf{X}_k(\omega))\Big(\sup_{\{\mathbf{x}:\, g(\mathbf{x})\ge g^*\}} m_g(\mathbf{x})\Big)^{n-n_\omega} \\ &\le \prod_{k=0}^{n_\omega} m_g(\mathbf{X}_k(\omega))\end{aligned}$$

Now, by (18.5), we have that the last term is almost surely finite. Therefore, according to (18.7), we deduce that $\limsup_{n\to\infty} g(\mathbf{X}_n)$ is almost surely finite, and taking into account (18.3), it is almost surely null, which concludes the proof.

Remark 18.3 Suppose, for example, the case $g(\mathbf{x}) = f$. Taking into account the interpretation of $m_g(\mathbf{x})$ given in Remark 18.2, assume that, for every $l \in \{1,\dots,N_m\}$, the family of ratios of couples per female, namely $\{L_l(\mathbf{x})/g(\mathbf{x})\}_{\mathbf{x}\in\mathbb{N}_+^2}$, is bound. Let $C_l := \sup_{\mathbf{x}\in\mathbb{N}_+^2} L_l(\mathbf{x})/g(\mathbf{x})$. Intuitively, Theorem 18.1 states that the extinction of the two-sex population happens if the average number of females per couple (for any reproductive strategy) do not exceed C_l^{-1}, that is

$$\max_{h\in\{1,\dots,N_r\}} g(\mu^h) \le \min_{l\in\{1,\dots,N_m\}} C_l^{-1}.$$

Similar interpretations can be given for the cases $g(\mathbf{x}) = m$ and $g(\mathbf{x}) = f + m$.

In order to investigate the possible survival of the population, given $g(\cdot)$ as in (18.1) and $\nu > 1$, we introduce the sequence of events $\{B_{k,\nu}\}_{k=0}^{\infty}$ where

$$B_{k,\nu} := \{g(\mathbf{X}_{k+1}) > \nu g(\mathbf{X}_k)\}. \tag{18.8}$$

In order to provide a result about the survival of the process, we previously establish the following auxiliar lemma.

Lemma 18.1 *Assume that there exists* $\nu > 1$ *such that*

$$\sum_{k=0}^{\infty} \sup_{\{\mathbf{x}:\, g(\mathbf{x})>\nu^k g(\mathbf{x}_0)\}} P(B_{k,\nu}^c \mid \mathbf{X}_k = \mathbf{x}) < \infty. \tag{18.9}$$

Then $P(\lim_{n\to\infty} g(\mathbf{X}_n) = \infty \mid \mathbf{X}_0 = \mathbf{x}_0) > 0.$

Proof

$$P(\lim_{n\to\infty} g(\mathbf{X}_n) = \infty \mid \mathbf{X}_0 = \mathbf{x}_0) \ge P(\cap_{k=0}^{\infty} B_{k,\nu} \mid \mathbf{X}_0 = \mathbf{x}_0)$$

$$\ge P(B_{0,\nu} \mid \mathbf{X}_0 = \mathbf{x}_0) \prod_{k=1}^{\infty} \inf_{\{\mathbf{x}:\, g(\mathbf{x})>\nu^k g(\mathbf{x}_0)\}} P(B_{k,\nu} \mid \mathbf{X}_k = \mathbf{x})$$

$$\ge P(B_{0,\nu} \mid \mathbf{X}_0 = \mathbf{x}_0) \prod_{k=1}^{\infty} (1 - \sup_{\{\mathbf{x}:\, g(\mathbf{x})>\nu^k g(\mathbf{x}_0)\}} P(B_{k,\nu}^c \mid \mathbf{X}_k = \mathbf{x})),$$

which is strictly positive taking into account that $P(B_{0,\nu} \mid \mathbf{X}_0 = \mathbf{x}_0) > 0$ and (18.9) holds.

Now we shall determine some conditions which ensure the survival of the two-sex population with a positive probability.

Theorem 18.2 *Let us consider the process* $\{\mathbf{X}_n\}_{n=0}^{\infty}$, $g(\cdot)$ *of the form (18.1) and* $g^* \geq 0$ *such that*

(i) $\inf_{\{\mathbf{x}:\, g(\mathbf{x})>g^*\}} m_g(\mathbf{x}) > 1$.
(ii) There exists $\gamma > 1$ *verifying that*

$$\sum_{k=0}^{\infty} \sup_{\{\mathbf{x}:\, g(\mathbf{x})>\nu^k g(\mathbf{x}_0)\}} \frac{M_g^{\gamma}(\mathbf{x})}{g(\mathbf{x})^{\gamma}} < \infty \tag{18.10}$$

where

$$M_g^{\gamma}(\mathbf{x}) := E\left[|g(\mathbf{X}_{k+1}) - E[g(\mathbf{X}_{k+1}) \mid \mathbf{X}_k = \mathbf{x}]|^{\gamma} \,|\, \mathbf{X}_k = \mathbf{x}\right].$$

Then, for all $\mathbf{x}_0 \in \mathbb{N}_+^2$ *with* $g(\mathbf{x}_0) > g^*$,

$$P(\lim_{n\to\infty} g(\mathbf{X}_n) = \infty \mid \mathbf{X}_0 = \mathbf{x}_0) > 0.$$

Proof From (a), there exist $\eta > 1$, $g(.)$ and $g^* > 0$ such that for all $\mathbf{x} \in \mathbb{N}_+^2$ with $g(\mathbf{x}) > g^*$, it is satisfied that $m_g(\mathbf{x}) \geq \eta > 1$. Consequently,

$$E\big[g(\mathbf{X}_{n+1}) \mid \mathbf{X}_n = \mathbf{x}\big] = m_g(\mathbf{x}) g(\mathbf{x}) \geq \eta g(\mathbf{x}). \tag{18.11}$$

Let $\varepsilon \in (0, \eta - 1)$ and $\nu := \eta - \varepsilon$. We consider the sequence $\{B_{k,\nu}\}_{k=0}^{\infty}$ as defined in (18.8). According to Lemma 18.1, it will be sufficient to verify (18.9).

Let $\mathbf{x} \in \mathbb{N}_+^2$ such that $g(\mathbf{x}) > g^*$. By using (18.11) and Markov's inequality, we get that

$$\begin{aligned}
P(B_{k,\nu}^c \mid \mathbf{X}_k = \mathbf{x}) &= P\left(g(\mathbf{X}_{k+1}) \leq (\eta - \varepsilon) g(\mathbf{x}) \mid \mathbf{X}_k = \mathbf{x}\right) \\
&\leq P\left(g(\mathbf{X}_{k+1}) \leq E[g(\mathbf{X}_{k+1}) \mid \mathbf{X}_k = \mathbf{x}] - \varepsilon g(\mathbf{x}) \mid \mathbf{X}_k = \mathbf{x}\right) \\
&\leq P\left(|g(\mathbf{X}_{k+1}) - E[g(\mathbf{X}_{k+1}) \mid \mathbf{X}_k = \mathbf{x}]| \geq \varepsilon g(\mathbf{x}) \mid \mathbf{X}_k = \mathbf{x}\right) \\
&\leq \frac{M_g^{\gamma}(\mathbf{x})}{\varepsilon^{\gamma} g(\mathbf{x})^{\gamma}}.
\end{aligned}$$

Now, applying (18.10), we deduce that (18.9) holds, which finishes the proof.

Remark 18.4 Condition (a) in Theorem 18.2 can be intuitively interpreted by a reasoning similar to the considered in Remark 18.3. Take, for example, $g(\mathbf{x}) = f$. Providing $D_l := \inf_{\mathbf{x}\in\mathbb{N}_+^2} L_l(\mathbf{x})/g(\mathbf{x}) > 0$, for every $l \in \{1, \ldots, N_m\}$, in order to guarantee a positive probability of survival it is sufficient an average number of females per couple greater (for any reproductive strategy) than D_l^{-1}, for every

$l \in \{1, \ldots, N_m\}$, i.e.

$$\min_{h \in \{1,\ldots,N_r\}} g(\mu^h) > \max_{l \in \{1,\ldots,N_m\}} D_l^{-1}.$$

With respect to the more technical condition (b) in Theorem 18.2, we remark here that it is not a very strong condition. In fact, it is satisfied, for example, if $M_g^{\gamma}(\mathbf{x}) = O(g(\mathbf{x}))$ as $g(\mathbf{x})$ tends to infinity. In such a case, it is easy to verify that the series in (18.10) is dominated by $\sum_{k=0}^{\infty} \nu^{-k(\gamma-1)}$ which is convergent.

18.5 Conclusions

Notice that for the female-male population-size dependent two-sex branching process studied in this work, the rate $m_g(\mathbf{x})$ plays a similar role in the extinction/survival of the population that the offspring mean for the classical Bienaymé-Galton-Watson branching process. Many questions remains for research about this new two-sex branching process, being necessary to complete both its probabilistic and its inferential theory. In particular, much of the research associated with two-sex branching processes has been concentrated on extinction probabilities. Significant progress has been made on this topic. However, research on the time to extinction has not been sufficiently investigated. This topic needs to be addressed with greater intensity. Assuming non-extinction, it is also necessary to investigate results about the limiting random variable. From an applied point of view, it is important to explore new fields where the new two-sex branching process can be applied as appropriate mathematical model, especially in semelparous species, which are characterized by a single reproductive episode before death, see [2] or [4].

Acknowledgements This research has been supported by the Ministerio de Economía y Competitividad of Spain (grants MTM2012-31235 and MTM2015-70522-P), the Junta de Extremadura (grant GR15105), the FEDER, and the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria (grant DFNI-I02/17).

References

1. Daley, D.J.: Extinction conditions for certain bisexual Galton–Watson branching processes. Z. Wahrscheinlichkeitsth. **9**, 315–322 (1968)
2. Fleming, I.A.: Reproductive strategies of Atlantic salmon: ecology and evolution. Rev. Fish Biol. Fish. **6**, 379–416 (2005)
3. Haccou, P., Jagers, P., Vatutin, V.: Branching Processes: Variation, Growth, and Extinction of Populations. Cambridge University Press, Cambridge (2005)
4. Hautekeete, N.C., Piquot, Y., Van Dijk, H.: Investment in survival and reproduction along a semelparity-iteroparity gradient in the Beta species complex. J. Evol. Biol. **14**, 795–804 (2001)
5. Kimmel, M., Axelrod, D.E.: Branching Processes in Biology. Springer, Berlin (2002)

6. Ma, S., Molina, M., Yongsheng, X.: Two-sex branching populations with progenitor couples in a random environment. Math. Popul. Stud. **19**, 177–187 (2012)
7. Molina, M., Mota, M., Ramos, A.: Bisexual Galton–Watson branching process with population–size–dependent mating. J. Appl. Probab. **39**, 479–490 (2002)
8. Molina, M., Jacob, C., Ramos, A.: Bisexual branching processes with offspring and mating depending on the number of couples in the population. Test **17**, 265–281 (2008)
9. Molina, M., Mota, M., Ramos, A.: Stochastic modeling in biological populations with sexual reproduction through branching models: application to Coho salmon populations. Math. Biosci. **258**, 182–188 (2014)
10. Xing, Y., Wang, Y.: On the extinction of a class of population-size dependent bisexual branching processes. J. Appl. Probab. **42**, 175–184 (2005)

Chapter 19
On Two-Type Decomposable Branching Processes in Continuous Time and Time to Escape Extinction

Maroussia Slavtchova-Bojkova

Mathematics Subject Classification (2000): 60J85, 60J80

19.1 Introduction

This paper is concerned with applications of branching models in different medical, biological and environmental contexts where we find a general pattern of populations that, due to a small reproductive ratio of the individuals/cells, will extinct after some time, but as a result of a random occurrence of mutations this trend could be changed dramatically. Such populations are, for example, the viruses which may become resistant after antibiotics treatment, some insects populations after hybridization and others. Our leading example will be the appearance of cancer cells after chemotherapy and we will now be interested in the most basic question regarding the evolutionary dynamics of cancer cells: how long does it take for a population to generate a single cell that will start a pathway with indefinite survival? Or in other words what is the probability of success or failure of the anti-cancer therapy? A typical situation of such populations is observed after chemotherapy (see e. g. Iwasa et al. [6, 7] and Nowak [10] together with references therein). The chemotherapy reduces the capacity of division of the cancer cells, which should lead to the destruction of tumors. However, sometimes mutations in the cells provide resistance to the therapy. This new type of cells has a higher reproduction and can avoid extinction.

M. Slavtchova-Bojkova (✉)
Department of Probability, Operations Research and Statistics, Faculty of Mathematics and Informatics, Sofia University, Sofia, Bulgaria

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
e-mail: bojkova@fmi.uni-sofia.bg

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3_19

Having in mind all the examples given above, it is of outstanding importance to have good estimates of the probability of escaping extinction and related aspects, such as the distribution of number of mutations which implies escaping extinction, the distribution of waiting times until escape. It was done in discrete-time setting by Serra and Haccou [12] and Serra [11] using discrete-time Galton–Watson branching processes (GWBP) and as pointed out there mathematically discrete-time models are much easier to handle than their continuous counterparts. In this paper we will generalize and expand some of these estimates in continuous-time setting using age-dependent branching processes. Although, that at first glance mathematically this seems a methodological step, it turns out to be not that easy to tackle such problems in a general setting.

Let us shortly remind that branching processes have been intensively studied during the last decades. Classical references are the books of Harris [5], Athreya and Ney [1], Jagers [8], and Mode [9]. For recent books, with emphasis on applications, see Axelrod and Kimmel [2], Haccou et al. [4] and also Durrett [3], especially for branching modeling in cancer. For a nice example of how branching processes can be used to solve important problems in biology and medicine, the reader is referred to the papers of Iwasa et al. [6, 7]. The close-related research related to the waiting times to extinction are some results of the author reviewed in the paper [13].

This paper is organized as follows: Sect. 19.2 introduces the branching process model with two types of cells in continuous time. Section 19.3 contains the main results and proofs. In Theorem 19.1 we prove the basic functional equations for probability generating function (p.g.f.) of the process itself. In Theorem 19.2 of Sect. 19.3.2 we obtain the p.g.f. of both the number of mutations occurred up to time t and the number of mutations to the escape type cells in the whole process. In the reminder of this section we studied the distribution of the waiting time T, which is actually the first moment in time when a mutation cell will start the lineage that will never go extinct (Theorem 19.3). More precisely we obtain its distribution, which is actually degenerate at infinity and the conditional expectation of T, conditioned of being finite. As a consequence of the results in Theorem 19.3 we show how one can obtain the probability of immediate escape from extinction in terms of modified hazard function of the random variable T, conditioned in addition on non-extinction of the process of type 1 cells, which have subcritical reproduction. Finally, we end with some concluding remarks and topics for further research.

19.2 Formulation of the Model

We will first outline an age-dependent branching process with one type of cells. Consider a cell proliferation process starting at time 0 with a single progenitor of type 1 of age 0 whose life-length τ has distribution $G(t) = P(\tau \leq t)$, $G(0^+) = 0$, i.e. $Z(0) = 1$. With probability p_k, $k \geq 0$, it produces at the end of its life k similar cells of age 0, which reproduce independently with the same distribution of the life-length τ and reproduction distribution $\{p_k, k \geq 0\}$, $\sum_{k=0}^{\infty} p_k = 1$. Provided that there is at least one offspring, the death-and-reproduction process is repeated, and continues as long as individuals/cells exist. The single-type process $\{Z(t), t \geq 0\}$ or the so-called Bellman–Harris branching process (BHBP) together with proper biological applications is studied by Jagers [8] and more theoretically by Athreya and Ney [1].

Now we present a two-type decomposable age-dependent branching model (also known as BHBP with two types of cells) $\{Z^0(t), Z^1(t), t \geq 0\}$, where $\{Z^0(t), t \geq 0\}$ and $\{Z^1(t), t \geq 0\}$ denote the number of cells of type 0 and type 1 at time t respectively. Suppose that cells of type 1 are subcritical, i.e. have reproduction mean m_1, $0 < m_1 < 1$, and that each one of their descendants can mutate at birth, independently of the others, to type 0 cells with probability u, $0 < u < 1$. Individuals/cells of type 0 are supercritical, i.e. have reproduction mean m_0, $1 < m_0 < \infty$, and there is no backward mutation. Let us mention here that if no mutation appear ($u = 0$) then the process will be described by two independent classical single-type BHBP.

By $G_i(t) = P(\tau_i \leq t)$, $G_i(0^+) = 0$, $i = 0, 1$ we denote the distribution of the life-lengths τ_i of the cells of type i, and by ν_i the offspring of type i cells, $i = 0, 1$.

Let us introduce the following notations:

$$F_i(t; s_0, s_1) = \mathbb{E}(s_0^{Z^0(t)} s_1^{Z^1(t)} | Z^i(0) = 1, Z^j(0) = 0, j \neq i), \quad \text{for} \quad i = 0, 1,$$
$$\mathbf{F}(t; \mathbf{s}) = (F_0(t; \mathbf{s}), F_1(t; \mathbf{s})),\ \mathbf{s} = (s_0, s_1).$$

Unless stated otherwise, we assume that the process starts with just one cell of type 1, i.e. $Z^0(0) = 0$ and $Z^1(0) = 1$. The p.g.f. of the offspring ν_i of type i cells will be denoted by $f_i(s)$, $i = 0, 1$.

Similar results of the discrete version of the two-type branching process, i.e. GWBP are obtained by Serra [11] and Haccou and Serra [12], where the distribution of the waiting time to produce a cell that will escape extinction, is studied.

19.3 Main Results

19.3.1 Basic Functional Equation

In the following theorem we will obtain the basic functional equation for the p.g.f. of the age-dependent branching process defined in Sect. 19.2.

Theorem 19.1 *The p.g.f.* $\mathbf{F}(t; s_0, s_1)$ *satisfies the following integral equations*

$$F_1(t; s_0, s_1) = s_1(1 - G_1(t)) + \int_0^t f_1(uF_0(t-y; s_0) + (1-u)F_1(t-y; s_0, s_1))dG_1(y), \tag{19.1}$$

and

$$F_0(t; s_0, s_1) \equiv F_0(t; s_0) = s_0(1 - G_0(t)) + \int_0^t f_0(F_0(t-y; s_0))dG_0(y), \tag{19.2}$$

where

$$F_i(0; s_0, s_1) = s_i, |s_i| \leq 1, i = 0, 1.$$

Proof We start with a derivation of the basic integral equation (19.1). A decomposition of the sample space Ω in accordance with the life-length τ_1 and number ν_1 of offspring of the initial cell of type 1 suggests the relation:

$$\begin{aligned}
F_1(t; s_0, s_1) &= \mathbb{E}\left(\mathbb{E}(s_0^{Z^0(t)} s_1^{Z^1(t)} | Z^1(0) = 1, Z^0(0) = 0, \{\tau_1, \nu_1\})\right) \\
&= s_1(1 - G_1(t)) \\
&+ \int_0^t dG_1(y) \sum_{j=0}^{\infty} p_{1j} \sum_{k=0}^{j} \binom{j}{k} \mathbb{E}(s_0^{Z^0(t-y)} u^k s_1^{Z^1(t-y)} (1-u)^{j-k} | Z^0(0) = k, Z^1(0) = j-k) \\
&= s_1(1 - G_1(t)) \\
&+ \int_0^t dG_1(y) \sum_{j=0}^{\infty} p_{1j} \sum_{k=0}^{j} \binom{j}{k} u^k (1-u)^{j-k} (\mathbb{E}(s_0^{Z^0(t-y)} | Z^0(0) = 1, Z^1(0) = 0))^k \times \\
&(\mathbb{E}(s_0^{Z^0(t-y)} s_1^{Z^1(t-y)} | Z^0(0) = 0, Z^1(0) = 1))^{j-k} \\
&= s_1(1 - G_1(t)) + \int_0^t f_1(uF_0(t-y; s_0) + (1-u)F_1(t-y; s_0, s_1))dG_1(y),
\end{aligned}$$

where $\{p_{1k}, k \geq 0\}$ is the distribution of the offspring of type 1 cells. Equation (19.2) is derived in a similar way. Notice that this equation is the integral equation obtained for the classical BHBP. □

Note that when G_1 is the unit step function

$$G_1(t) = \begin{cases} 0, \text{ for } t < 1, \\ 1, \text{ for } t \geq 1, \end{cases}$$

then (19.1) reduces to a functional iteration formula for $F(n; s_0, s_1)$ obtained by Serra in [11]; while if

$$G_1(t) = \begin{cases} 0, & \text{for } t < 0, \\ 1 - e^{-\lambda t}, & \text{for } t \geq 0, \end{cases}$$

then we have a two-type Markov branching process allowing mutations.

Let us mention here that the result in continuous time is rather different from that of Serra [11] using GWBP, where actually the p.g.f. of the process is reduced to the single-type GWBP and after that is used significantly to study the distribution of the number of mutations. On the other hand, here we would like to point out that using Eqs. (19.1) and (19.2) one can study the asymptotic properties of the mean, variance and higher moments of types 0 and 1 cells when $t \to \infty$, which is left for a later stage.

19.3.2 Number of Mutants and Probability of Extinction

Unless mutations occur, the process of interest will be a single-type subcritical BHBP and it is the appearance of mutants that makes the study of such populations an interesting task. That is why it is important to study the total number of mutations that occur in the whole process. This random quantity will play a crucial role in determining the extinction probability of the process.

Consider the random variable (r.v.) $I(t)$, $t \geq 0$, being the total number of mutants produced until time t (inclusive), and let I be the r.v. that represents the number of mutants in the whole process. By mutant we mean a cell of type 0, whose mother is of type 1. It is obvious that the sequence of r.v. $I(t), t \geq 0$, converges pointwise to the r. v. I. In our next theorem, we use this convergence to establish a functional equation for the p.g.f. of I, denoted by $h_I(s)$.

Theorem 19.2 *The p.g.f.* $h_I(s)$ *of* I *and* $h_{I(t)}(s)$ *of* $I(t)$ *satisfy the functional equations*

$$h_I(s) = f_1(us + (1-u)h_I(s)), \tag{19.3}$$

$$h_{I(t)}(s) = 1 - G_1(t) + \int_0^t f_1(us + (1-u)h_{I(t-y)}(s))dG_1(y), \tag{19.4}$$

for all $s \in [0, 1]$.

Proof First we establish a recursive relation for the p.g.f. of the r.v. $I(t), t \geq 0$. We will use again a decomposition of the sample space Ω in accordance with the life-length τ_1 and the number of offspring ν_1 of the initial cell of type 1. It is clear that $I(t) = 0$, if $\tau_1 > t$ with probability $1 - G_1(t)$, and $I(t) = \nu_{10} + \sum_{i=1}^{\nu_1 - \nu_{10}} I(t - \tau_1)$, when $\tau_1 \leq t$, where $\nu_{10} \in Bi(\nu_1, u)$ is the number of mutations between the descendants and $\nu_1 - \nu_{10}$ is the number of type 1 cells, produced by the initial cell. Therefore

$$
\begin{aligned}
&\mathbb{E}(s^{I(t)}|Z^1(0) = 1, Z^0(0) = 0)\\
&= (1 - G_1(t))s^0 + \int_0^t dG_1(y)\mathbb{E}\left(s^{\nu_{10} + \sum_{i=0}^{\nu_1 - \nu_{10}} I_i(t-y)} \right)\\
&= 1 - G_1(t) + \int_0^t dG_1(y)\mathbb{E}\left(s^{\nu_{10}} \times s^{\sum_{i=0}^{\nu_1 - \nu_{10}} I_i(t-y)} \right)\\
&= 1 - G_1(t) + \int_0^t dG_1(y)\mathbb{E}\left[\mathbb{E}\left(s^{\nu_{10}} \times s^{\sum_{i=0}^{\nu_1 - \nu_{10}} I_i(t-y)} \right) |(\nu_1, \nu_{10}) \right]\\
&= 1 - G_1(t) + \int_0^t \left[\sum_{j=0}^{\infty} p_{1j} \sum_{k=0}^{j} \binom{j}{k} u^k (1-u)^{j-k} s^k \times \mathbb{E}s^{\sum_{i=0}^{j-k} I_i(t-y)} \right] dG_1(y)\\
&= 1 - G_1(t) + \int_0^t \left[\sum_{j=0}^{\infty} p_{1j} \sum_{k=0}^{j} \binom{j}{k} u^k (1-u)^{j-k} s^k \times (h_{I(t-y)}(s))^{j-k} \right] dG_1(y)\\
&= 1 - G_1(t) + \int_0^t f_1(us + (1-u)h_{I(t-y)}(s))dG_1(y),
\end{aligned}
$$

where $I_i(t)$ are independent identically distributed copies of $I(t)$ and $h_{I(0)}(s) = 1$. After that, using the techniques of embedded generation process (see Athreya and Ney, p. 141 [1]) we obtain that the result of the number of mutations in the whole process proved by Serra [11] for Galton–Watson branching processes remains true for age-dependent branching processes. □

We now proceed to determine the probability of extinction. Using the notation

$$
\begin{aligned}
q_0 &= \mathbf{P}[Z^0(t) = Z^1(t) = 0 \ \text{ for some } \ t > 0|Z^0(0) = 1, Z^1(0) = 0],\\
q_1 &= \mathbf{P}[Z^0(t) = Z^1(t) = 0 \ \text{ for some } \ t > 0|Z^0(0) = 0, Z^1(0) = 1],
\end{aligned}
$$

it follows, from the classical result on the extinction of branching processes, that q_0 is the smallest root of the equation $q_0 = f_0(q_0)$ in the interval $[0, 1]$. To determine q_1, notice that extinction of the process occurs if and only if all the supercritical single-type BHBP starting from the mutants die out, since $m_1 < 1$. Therefore, since there are I such processes, we have

$$q_1 = \mathbf{E}[q_0^I] = h_I(q_0).$$

Then, we deduce that $q_1 < 1$, since $m_0 > 1$ and $q_0 < 1$. Let us remind that assuming small mutation rate u, Iwasa et al. [6, 7] provided approximations for particular reproduction laws, namely for Poisson and geometric distributions. Their results extend to even more complex scheme of mutations leading to branching processes with more than two types of individuals.

19.3.3 Time to Escape Extinction

Now, we consider the r. v. T, which represents the time to escape extinction, i.e. the first time in which a successful mutant is produced. By successful mutant we mean a mutant that is able to start a single-type BHBP that allows indefinite survival. This variable takes values in the set $(0, +\infty]$, with $T = \infty$, if no successful mutant is produced.

Theorem 19.3 *The distribution of T has the following properties:*

(i) $\mathbb{P}(T > t) = h_{I(t)}(q_0) \equiv Q_t, \quad for \quad t > 0,$

(ii) $\mathbb{P}(T = \infty) = q_1,$

(iii) $\mathbb{E}(T|T < \infty) = \dfrac{1}{1-q_1}\displaystyle\int_0^\infty [h_{I(t)}(q_0) - q_1]dt,$

where Q_t are defined by

$$Q_t = 1 - G_1(t) + \int_0^t f_1(uq_0 + (1-u)Q_{t-y})dG_1(y)$$

with $Q_0 = 1$.

Proof To prove (i), observe that $T > t$ means that all mutants that occurred up to time t were unsuccessful. Therefore,

$$\mathbb{P}(T > t) = \mathbb{E}(q_0^{I(t)}) = h_{I(t)}(q_0).$$

To prove (ii), observe that $(T > t)_{t\geq 0} \downarrow (T = \infty)$ and it implies

$$\mathbb{P}(T > t)_{t\geq 0} \downarrow \mathbb{P}(T = \infty).$$

Then from (i)

$$\lim_{t\to\infty} \mathbb{P}(T > t) = \lim_{t\to\infty} h_{I(t)}(q_0)$$

and as $I(t)$ converges pointwise to I, it follows

$$\mathbb{P}(T = \infty) = \lim_{t \to \infty} h_{I(t)}(q_0) = h_I(q_0) = q_1.$$

For proving (iii) observe that $T > 0$ and, therefore,

$$\begin{aligned}
\mathbb{E}(T|T < \infty) &= \int_0^\infty (1 - \mathbb{P}(T \leq t|T < \infty))dt \\
&= \int_0^\infty (1 - \frac{\mathbb{P}(T \leq t, T < \infty)}{1 - q_1})dt \\
&= \frac{1}{1 - q_1} \int_0^\infty (1 - q_1 - \mathbb{P}(T \leq t))dt \\
&= \frac{1}{1 - q_1} \int_0^\infty (\mathbb{P}(T > t) - q_1)dt \\
&= \frac{1}{1 - q_1} \int_0^\infty (h_{I(t)}(q_0) - q_1)dt,
\end{aligned}$$

with $h_{I(t)}(s)$ defined by equation (19.4). □

19.3.4 Immediate Risk of Escape

Another natural characterization of the appearance of a successful mutant is the probability of producing a successful mutant in a very short time interval dt after time t, given that it has not been produced yet, called immediate risk of escape extinction. We will show in this subsection how one can compute this probability theoretically using the results of Theorem 19.3.

In general one could use the hazard function of the variable T, defined by $\mathbb{P}(T \in (t, t + dt)|T > t)$, but in this case we need to modify this function, as it is done in discrete-time setting (see Serra and Haccou [12]). It is due to the fact that the r. v. T has a defective distribution ($T = \infty$ when no successful mutant is produced) and in fact, if there are no subcritical individuals (i.e. of type 1) alive at time t, the probability of producing a successful mutant immediately after this moment is zero. That is why we will use the following modification of the standard hazard function:

$$g(t)dt = \mathbb{P}(T \in (t, t + dt|T > t, Z^1(t) > 0)). \tag{19.5}$$

We have

$$g(t)dt = \frac{\mathbb{P}(T \in (t, t+dt))}{\mathbb{P}(T > t, Z^1(t) > 0)}$$
$$= \frac{\mathbb{P}(T \in (t, t+dt))}{\mathbb{P}(T > t) - \mathbb{P}(T > t, Z^1(t) = 0)}.$$

The probabilities $\mathbb{P}(T > t)$ and $\mathbb{P}(T \in (t, t+dt))$ are computed using Theorem 19.3. The second term in the denominator satisfies the following recursive formula:

$$\mathbb{P}(T > t, Z^1(t) = 0) = \mathbb{P}(T > t, Z^1(t) = 0 | \tau_1 < t)\mathbb{P}(\tau_1 < t)$$
$$+ \mathbb{P}(T > t, Z^1(t) = 0 | \tau_1 \geq t)\mathbb{P}(\tau_1 \geq t)$$
$$= \int_0^t \mathbb{P}(T > t, Z^1(t) = 0 | \tau_1 = y) dG_1(y)$$
$$= \int_0^t \sum_{j=0}^{\infty} p_{1j} \sum_{k=0}^{j} \binom{j}{k} (uq_0)^k \left((1-u)\mathbb{P}(T > t-y, Z^1(t-y) = 0)\right)^{j-k} dG_1(y)$$
$$= \int_0^t f_1(uq_0 + (1-u)\mathbb{P}(T > t-y, Z^1(t-y) = 0)) dG_1(y),$$

using that τ_1 is the life-time and $\{p_{1k}, k \geq 0\}$ is the distribution of the offspring of the initial cell of type 1 and clearly the second conditional probability term is equal to 0.

Let us remind here that we receive the similar recursive formula as that established by Serra and Haccou [12].

Indeed, from

$$\mathbb{P}(T > t, Z^1(t) = 0) = \int_0^t f_1(uq_0 + (1-u)\mathbb{P}(T > t-y, Z^1(t-y) = 0)) dG_1(y) \tag{19.6}$$

when

$$G_1(t) = \begin{cases} 0, \text{ for } t < 1, \\ 1, \text{ for } t \geq 1, \end{cases}$$

we obtain that

$$\mathbb{P}(T > n, Z^1(n) = 0) = f_1(uq_0 + (1-u)\mathbb{P}(T > n-1, Z^1(n-1) = 0)), \quad n \geq 1,$$

which is exactly the result in [12].

So, the modified hazard function is given by

$$
\begin{aligned}
g(t)dt &= \frac{\mathbb{P}(T \in (t, t+dt))}{\mathbb{P}(T > t) - \mathbb{P}(T > t, Z^1(t) = 0)} \\
&= \frac{Q_t' dt}{Q_t - \mathbb{P}(T > t, Z^1(t) = 0)}
\end{aligned}
$$

where $\mathbb{P}(T > t, Z^1(t) = 0)$ satisfies Eq. (19.6).

19.4 Concluding Remarks

First, we would like to conclude that this paper is the first step towards the expanding of the theory and methods using continuous-time counterparts of the discrete-time GWBP for different schemes leading to mutations. In the context of cancer dynamics, resistance to the anti-cancer therapy and the possible appearance of metastasis we are tackling the problems of the distribution of the first moment of the occurrence of "successful" mutant together with the growth of the population of the mutant cells. Let us mention here that once the results are proved for models with one type of mutation, they could be extended to more than one type and different mutation schemes, including backward and/or forward mutations. Moreover, the branching models with continuous-time are more realistic and reveal more adequately and accurately the behavior of cell populations with overlapping generations.

Secondly, we found the analytical decisions in terms of p.g.f. of these r.v.s, which could be used subsequently to derive their moments and corresponding limit theorems for the BHBP driving the development of the real process, which might be considered per se as an innovation in application of branching theory. In this connection, the following questions for further research could be pointed out: to study the distribution of the waiting time to attain certain levels of the branching processes in continuous-time setting and to obtain limiting results for its distribution.

Thirdly, we obtain also a theoretical formulae for immediate risk of avoiding extinction that could be used later on for the comparison of the behavior of the modified hazard function for different offspring distributions.

Finally, the results proved generalize the similar ones obtained by Serra [11], Serra and Haccou [12] and those for the exponential models of growth developed by Iwasa et al. [6, 7] and Durrett [3], as well.

Acknowledgements The research is partially supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No DFNI-I02/17 and by Ministerio de Educación y Ciencia and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grant MTM2012-31235. The author is grateful to Prof. Kosto Mitov for his valuable support and helpful discussion in preparing the manuscript and to the antonymous referee for the detailed remarks, advices and comments, which lead to the significant improvement of the paper.

References

1. Athreya, K.B., Ney, P.: Branching Processes. Springer, Berlin (1972)
2. Axelrod, D.E., Kimmel, M.: Branching Processes in Biology. Springer, New York (2002)
3. Durrett, R.: Branching Process Models in Cancer. Springer, Cham (2015)
4. Haccou, P., Jagers, P., Vatutin, V.: Branching Processes: Variation, Growth and Extinction of Populations. Cambridge University Press, Cambridge (2005)
5. Harris, T.E.: The Theory of Branching Processes. Springer, Berlin (1963)
6. Iwasa, Y., Michor, F., Nowak, M.: Evolutionary dynamics of escape from biomedical intervention. Proc. Biol. Sci. Lond. B **270**, 2573–2578 (2003)
7. Iwasa, Y., Michor, F., Nowak, M.: Evolutionary dynamics of invasion and escape. J. Theor. Biol. **226**, 205–214 (2004)
8. Jagers, P.: Branching Processes with Biological Applications. Wiley, New York (1975)
9. Mode, C.J.: Multitype Branching Processes: Theory and Applications. Elsevier, New York (1971)
10. Nowak, M.A., Michor, F., Komarova, N.L., Iwasa, Y.: Evolutionary dynamics of tumor suppressor gene inactivation. Proc. Natl. Acad. Sci. U. S. A. **101**, 10635–10638 (2004)
11. Serra, M.: On waiting time to escape. J. Appl. Probab. **43**, 296–302 (2006)
12. Serra, M., Haccou, P.: Dynamics of escape mutants. Theor. Popul. Biol. **72**, 167–178 (2007)
13. Slavtchova-Bojkova, M.: Time to extinction in branching processes and its application in epidemiology. Pliska Stud. Math. Bulg. **22**, 101–126 (2013)

Index

I.M. del Puerto et al. (eds.), *Branching Processes and Their Applications*, Lecture Notes in Statistics 219, DOI 10.1007/978-3-319-31641-3