

# Lecture Notes in Statistics – Proceedings

197

Edited by P. Bickel, P. Diggle, S. Fienberg, U. Gather,  
I. Olkin, S. Zeger

For further volumes:  
<http://www.springer.com/series/694>



Miguel González · Inés M. del Puerto ·  
Rodrigo Martínez · Manuel Molina · Manuel Mota ·  
Alfonso Ramos  
Editors

# Workshop on Branching Processes and Their Applications

 Springer

*Editors*

Dr. Miguel González  
Departamento de Matemáticas  
Facultad de Ciencias  
Universidad de Extremadura  
Avda. Elvas, s/n.  
06006-Badajoz  
Spain  
mvelasco@unex.es

Dr. Inés M. del Puerto  
Departamento de Matemáticas  
Facultad de Ciencias  
Universidad de Extremadura  
Avda. Elvas, s/n.  
06006-Badajoz  
Spain  
idelpuerto@unex.es

Dr. Rodrigo Martínez  
Centro Universitario de Plasencia  
Universidad de Extremadura  
Avda. Virgen del Puerto, s/n.  
10600-Plasencia  
Spain  
rmartinez@unex.es

Dr. Manuel Molina  
Departamento de Matemáticas  
Facultad de Ciencias  
Universidad de Extremadura  
Avda. Elvas, s/n.  
06006-Badajoz  
Spain  
mmolina@unex.es

Dr. Manuel Mota  
Departamento de Matemáticas  
Facultad de Ciencias  
Universidad de Extremadura  
Avda. Elvas, s/n.  
06006-Badajoz  
Spain  
mota@unex.es

Dr. Alfonso Ramos  
Facultad de Veterinaria  
Avda. de la Universidad, s/n.  
10003-Cáceres  
Spain  
aramos@unex.es

ISSN 0930-0325

ISBN 978-3-642-11154-9

e-ISBN 978-3-642-11156-3

DOI 10.1007/978-3-642-11156-3

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010920305

© Springer-Verlag Berlin Heidelberg 2010

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Cover design:* Integra Software Services Pvt. Ltd., Pondicherry

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Foreword

One of the charms of mathematics is the contrast between its generality and its applicability to concrete, even everyday, problems.

Branching processes are typical in this. Their niche of mathematics is the abstract pattern of reproduction, sets of individuals changing size and composition through their members reproducing; in other words, what Plato might have called the pure idea behind demography, population biology, cell kinetics, molecular replication, or nuclear fission, had he known these scientific fields. Even in the performance of algorithms for sorting and classification there is an inkling of the same pattern.

In special cases, general properties of the abstract ideal then interact with the physical or biological or whatever properties at hand. But the population, or branching, pattern is strong; it tends to dominate, and here lies the reason for the extreme usefulness of branching processes in diverse applications. Branching is a clean and beautiful mathematical pattern, with an intellectually challenging intrinsic structure, and it pervades the phenomena it underlies.

The problem that gave birth to branching processes, that of the astonishingly frequent extinction of noble families, illustrates the situation well. The starting point was a question, which today would seem more intriguing or amusing than serious: Where have all the well-known families gone? Where are all the Hohenstaufers, Plantagenets, Tudors, Trastámaras, or Vasas? How come that they so often have died out, even though the populations they were part of increased? Degeneration?

But to Bienaymé and Galton, 150 years ago, this was a serious concern, and behind it lurks a question of general scientific interest: Can frequent extinction of separate family lines go hand in hand with a rapid increase of the whole population? Is this a general property of the branching pattern and not some oddity of human populations? If so, would it be part of an explanation of the extremely frequent extinction of species?

After a first disastrous mistake and 50 years of ensuing confusion, branching processes were able to prove that in natural conditions frequent extinction of separate family lines is completely compatible with exponential (“Malthusian”) growth of

the whole. This is, indeed, a mathematical property: a strongly supercritical branching process can still exhibit extinction probabilities say around 75%.

Extinction matters, like the time and path to extinction, remain important in branching processes and cutting-edge areas of application like conservation biology. But with time branching processes have developed into a full-fledged theory of population dynamics, encompassing the growth of populations as well as the stabilization of their composition, and their pedigrees looking backward. Mating can be dealt with, at least in simple processes, the effects of random environments clarified, and the first steps are being taken on models where there is a feedback from population size on individual behaviour. Today branching processes is a mature and important part of mathematics, and also of theoretical biology, theoretical chemistry, nuclear physics, computer algorithms, and demography.

Many aspects of the theory and its stance today were presented at the 2009 Badajoz workshop on branching processes, organized by the active branching process group of the University of Extremadura. This book contains most of the lectures given. It will be of great help to those wanting to acquaint themselves with contemporary branching process theory.

Gothenburg, Sweden  
September 2009

*Peter Jagers*

# Preface

The Workshop on Branching Processes and their Applications (WBPA) was held during 20–23 April 2009 in Badajoz, Spain. This conference gave continuity to such important previous meetings as the First World Conference on Branching Processes held in Varna, Bulgaria, in 1993, the IMA workshop on Classical and Modern Branching Processes held in Minnesota, USA, in 1994, and the more recent Symposium: Branching in Biology held in Gothenburg, Sweden, in 2005. The WBPA was promoted and organized by the Branching Processes Research Group belonging to the Department of Mathematics of the University of Extremadura, Spain. There were 35 invited participants from 15 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. They clearly indicated the importance of branching concepts in the development of theoretical approaches to solving problems in applied fields such as Epidemiology, Cell Kinetics, Genetics, and, of course Population Dynamics.

The Proceedings consists of 20 papers. All of them have been thoroughly reviewed. Parts covered by the workshop have been classified into the following areas:

1. Population Growth Models in Random and Varying Environments
2. Special Branching Processes
3. Limit Theorems and Statistics
4. Applications in Cell Kinetics and Genetics
5. Applications in Epidemiology
6. Two-Sex Branching Models

The first part deals with Population Growth Models in Random and Varying Environments. V. Vatutin considers critical branching processes in independent and identically distributed random environment. He shows the asymptotic behaviour of the survival probability and proves a conditional functional limit theorem under hypotheses which are milder than those used in classical papers. G. Alsmeyer revisits the extinction problem in branching processes in a stationary ergodic environment.

The use of random times in connection with the stationary environment leads him quite naturally to the use of Palm-duality theory in some of his arguments. C. Braumann compares the density-independent models of population growth; namely the Galton–Watson process, the simple birth and death process, and the Malthusian stochastic differential equation model, the first two being demographic stochasticity models, and the third an environmental stochasticity model. P. Mayster establishes the existence of stationary distributions for alternating branching processes, where two Markov branching processes act alternately in random time periods of observation and treatment.

In Part 2, Special Branching Processes, F. Klebaner considers models of population–size–dependent branching processes with the feature that they are supercritical when the population reaches a certain threshold, near critical around that value, and subcritical below it. G. Yanev reviews the existing results and presents new ones on certain subtrees of the Galton–Watson family tree. He considers rooted and complete subtrees, i.e., subtrees rooted at the ancestor and being family trees of a deterministic branching process. K. Mitov et al. study Bienaymé–Galton–Watson processes subordinated by a renewal process for which the interarrival periods have a finite mean or heavy tails. V. Topchii studies renewal measure densities associated with the problem of determining the expected number of particles at the origin of catalytic branching random walks.

Part 3 focuses on Limit Theorems and Statistics. I. Rahimov considers a branching stochastic process with non-stationary immigration given by an offspring distribution depending on an unknown parameter. He estimates this parameter and introduces a bootstrap process. The paper deals with how good the estimator must be in order for the bootstrap process to have the same asymptotic properties as the original process. M. Ispány and G. Pap investigate critical and nearly critical Galton–Watson branching processes with immigration, obtaining related functional limit theorems by using a general convergence theorem for martingale differences. M. González and I. del Puerto propose a weighted conditional least squares estimator of the offspring mean matrix for a multitype controlled branching process and study its asymptotic properties in the supercritical case.

Part 4 comprises some applications of the branching processes theory in Cell Kinetics and Genetics. N. Yanev considers some new ideas for branching process theory that arise in cell proliferation modeling. He considers distributions of discrete and continuous labels and of ages and residual lifetimes, models of leukemia cell kinetics, age-dependent branching populations with randomly chosen paths of evolution as models of (in vitro) progenitor cell populations and the estimation of offspring distributions, multitype branching populations with a large number of ancestors, and asymptotic likelihood estimation of the basic mitotic parameters. M. Kimmel and M. Mathaes propose a modification of the discrete time branching process described by Griffith and Pakes to model the amplification, mutation, and selection forces of Alu elements. They derive a limit frequency spectrum of the Alu element distribution, which serves as the theoretical, neutral frequency with which



real Alu insertion data can be compared through statistical goodness of fit tests. M. González et al. use a two-type bisexual branching process to model, in a two-sex monogamic population, the evolution of the number of carriers of the two alleles of a gene linked to the *Y* chromosome. They deal with inferential problems arising from this model, considering a frequentist and parametric approach. They consider the situation in which the only data available are the total number of females and the total number of males of each genotype in each generation. This leads them to use the expectation-maximization (EM) method in order to obtain maximum likelihood estimators.

Part 5 is about applications in Epidemiology. This is an applied area in which a number of new and exciting contributions were made at the Workshop. F. Ball and P. Neal are concerned with applications of branching processes to model the spread of an SIR (susceptible–infective–removed) epidemic among a closed, homogeneously mixing population consisting initially of certain numbers of infective and susceptible individuals. Each infective remains infectious for a period sampled independently from an arbitrary but specified distribution, during which he/she contacts susceptible individuals independently with some rate for each susceptible. C. Jacob et al. treat the problem of modeling the propagation of Bovine Spongiform Encephalopathy at the scale of a very large population (Great Britain) in order to predict its extinction time and to evaluate the efficiency of the main feed-ban regulation. They elaborate a multitype branching process in discrete time with age and population dependent individual transitions. The types are the health states at each age. M. González et al. use a Sevast’yanov age-dependent branching process to describe outbreaks of infectious diseases with an incubation period. They propose a method to obtain the optimal proportion of susceptible individuals that have to be vaccinated in order to eliminate the disease from the population. D. Heinzmann is interested in modeling the transmission dynamics of the macroparasite *Echinococcus granulosus*. He presents an approximation for the time to extinction in a sub-critical epidemic two-host interaction process for this macroparasite by using multitype branching processes.

Part 6, Two-Sex Branching Models, has contributions from S. Ma and Y. Xing who introduce and study a class of discrete time bisexual branching processes in which in each generation there is allowed the immigration of females and males, depending on the current numbers of females and males in the population, and from M. Molina who presents a summary of the literature associated with the classes of two-sex branching processes.

The organizers greatly appreciate the major response from the participants to submit contributions to the Proceedings. 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 Spanish Ministry of Science and Innovation (Ministerio de Ciencia e Innovación), the University of Extremadura itself, and the Local Administration. The Spanish Society of Statistics and Operations Research (SEIO) supported the meeting scientifically.

Also, our very special thanks go to Peter Jagers for accepting the task of writing the foreword of this book. Finally, many thanks to Springer Publishers, particularly to Niels Peter Thomas and John Kimmel, for making it possible for this project to see the light.

Badajoz, Spain  
September 2009

*Miguel González*  
On behalf of the Editors

# Contents

## Part I Population Growth Models in Random and Varying Environments

<b>1</b>	<b>A refinement of limit theorems for the critical branching processes in random environment</b> . . . . .	3
	Vladimir Vatutin	
1.1	Introduction and main results . . . . .	3
1.2	Branching in conditioned environment . . . . .	8
1.3	Proof of Theorems 1.1 and 1.2 . . . . .	17
	References . . . . .	18
<b>2</b>	<b>Branching processes in stationary random environment: The extinction problem revisited</b> . . . . .	21
	Gerold Alsmeyer	
2.1	Introduction . . . . .	21
2.2	Classical results revisited . . . . .	24
2.3	Main result and a counterexample . . . . .	25
2.4	Some useful facts from Palm-duality theory . . . . .	28
2.5	Proofs . . . . .	29
	References . . . . .	36
<b>3</b>	<b>Environmental versus demographic stochasticity in population growth</b> . . . . .	37
	Carlos A. Braumann	
3.1	Introduction . . . . .	37
3.2	Density-independent models and their local behavior . . . . .	39
3.3	Density-independent models and extinction . . . . .	44
3.4	Density-dependent models for environmental stochasticity . . . . .	45
3.5	Conclusions . . . . .	49
	References . . . . .	52

**4 Stationary distributions of the alternating branching processes . . . . . 53**  
 Penka Mayster

- 4.1 Introduction . . . . . 53
- 4.2 Alternating branching process . . . . . 55
- 4.3 Alternating branching process with explicit immigration . . . . . 56
- 4.4 Reproduction by  $n$  cycles . . . . . 58
- 4.5 Criticality . . . . . 59
- 4.6 Stationary distribution in random environment . . . . . 62
- 4.7 Unconditional probability generating functions . . . . . 63
- 4.8 Feed-back control . . . . . 64
- References . . . . . 66

**Part II Special Branching Processes**

**5 Approximations in population-dependent branching processes . . . . . 71**  
 Fima C. Klebaner

- 5.1 Introduction and a motivating example . . . . . 71
- 5.2 A Representation of the process and its re-scaled version . . . . . 73
  - 5.2.1 Re-scaled process: Dynamics plus small noise . . . . . 74
  - 5.2.2 Dynamics without noise in binary splitting . . . . . 74
- 5.3 Time to extinction . . . . . 75
- 5.4 The size of the population after a long time provided  
 it has survived . . . . . 75
- 5.5 Case of small initial population . . . . . 76
  - 5.5.1 Probability of becoming large and time for it to happen . . 77
- 5.6 Behaviour before extinction . . . . . 77
- References . . . . . 78

**6 Extension of the problem of extinction on Galton–Watson family  
 trees . . . . . 79**  
 George P. Yanev

- 6.1 Introduction . . . . . 79
- 6.2 Critical phenomenon . . . . . 80
- 6.3 Distribution of the number of complete and disjoint subtrees,  
 rooted at the ancestor . . . . . 83
- 6.4 Ratio of expected values of  $Z_n$ s provided infinite subtrees exist . . . 84
- 6.5 Geometric offspring distribution . . . . . 86
- 6.6 Poisson offspring distribution . . . . . 89
- 6.7 One-or-many offspring distribution . . . . . 91
- 6.8 Concluding remarks . . . . . 93
- References . . . . . 93

**7 Limit theorems for critical randomly indexed branching processes . . 95**  
 Kostov V. Mitov, Georgi K. Mitov and Nikolay M. Yanev

- 7.1 Introduction . . . . . 95
- 7.2 A conditional limit theorem for random time change . . . . . 97

- 7.3 Renewal processes ..... 100
- 7.4 BGW branching processes starting with random number  
of particles ..... 103
- 7.5 Limit theorems for the process  $Y(t)$  ..... 105
- 7.6 Concluding remarks ..... 107
- References ..... 108
- 8 Renewal measure density for distributions with regularly varying  
tails of order  $\alpha \in (0, 1/2]$  ..... 109**  
Valentin Topchii
- 8.1 Introduction ..... 109
- 8.2 Effects of attraction to a stable law ..... 111
- 8.3 Asymptotics of renewal function density ..... 114
- References ..... 118

**Part III Limit Theorems and Statistics**

- 9 Approximation of a sum of martingale differences generated by a  
bootstrap branching process ..... 121**  
Ibrahim Rahimov
- 9.1 Introduction ..... 121
- 9.2 Main theorems ..... 123
- 9.3 Array of processes ..... 127
- References ..... 133
- 10 Critical branching processes with immigration ..... 135**  
Márton Ispány and Gyula Pap
- 10.1 Introduction ..... 135
- 10.2 Branching and autoregressive processes ..... 136
- 10.3 Functional limit theorems ..... 138
- 10.4 Nearly critical branching processes with immigration ..... 140
- 10.5 Conditional least squares estimators ..... 142
- References ..... 145
- 11 Weighted conditional least squares estimation in controlled  
multitype branching processes ..... 147**  
Miguel González and Inés M. del Puerto
- 11.1 Introduction ..... 147
- 11.2 Probability model ..... 148
- 11.3 Weighted conditional least squares estimator of the offspring  
mean matrix ..... 150
- References ..... 155

## Part IV Applications in Cell Kinetics and Genetics

<b>12</b>	<b>Branching processes in cell proliferation kinetics</b> . . . . .	159
	Nikolay M. Yanev	
12.1	Introduction . . . . .	159
12.2	Distributions of discrete marks over a proliferating cell populations . . . . .	161
12.3	Distributions of continuous labels in branching populations of cells . . . . .	162
12.4	Age and residual lifetime distributions for branching processes . . .	164
12.5	Branching processes with immigration as models of leukemia cell kinetics . . . . .	168
12.6	Age-dependent branching populations with randomly chosen paths of evolution . . . . .	171
12.7	Multitype branching populations with a large number of ancestors . . . . .	173
12.8	Concluding remarks . . . . .	177
	References . . . . .	177
<b>13</b>	<b>Griffiths–Pakes branching process as a model for evolution of Alu elements</b> . . . . .	179
	Marek Kimmel and Matthias Mathaes	
13.1	Introduction . . . . .	179
13.2	Alu repeat sequences . . . . .	180
	13.2.1 Background on Alus . . . . .	180
	13.2.2 Alu sequence data used in this study . . . . .	180
13.3	Discrete branching process of Griffiths and Pakes with infinite allele mutations . . . . .	181
	13.3.1 Linear fractional offspring distribution . . . . .	184
13.4	Fitting results . . . . .	186
13.5	Discussion . . . . .	187
	References . . . . .	189
<b>14</b>	<b>Parametric inference for Y-linked gene branching models: Expectation-maximization method</b> . . . . .	191
	Miguel González, Cristina Gutiérrez and Rodrigo Martínez	
14.1	Introduction . . . . .	191
14.2	The probability model . . . . .	192
14.3	The estimation problem: The expectation-maximization method . .	195
	14.3.1 Determining the distribution of $\mathcal{F}Rr_N   (\mathcal{F}M_N, \alpha, \theta_R, \theta_r)$ . . . . .	197
	14.3.2 The expectation-maximization method . . . . .	199
14.4	Simulation study . . . . .	200
	References . . . . .	204

**Part V Applications in Epidemiology**

**15 Applications of branching processes to the final size of SIR epidemics** . . . . . 207  
 Frank Ball and Peter Neal

15.1 Introduction . . . . . 208

15.2 Early stages of epidemic . . . . . 211

15.3 Final outcome of Reed–Frost epidemic . . . . . 213

    15.3.1 Preliminaries . . . . . 213

    15.3.2 Many initial infectives . . . . . 216

    15.3.3 Few initial infectives . . . . . 219

    15.3.4 Central limit theorem . . . . . 221

References . . . . . 222

**16 A branching process approach for the propagation of the Bovine Spongiform Encephalopathy in Great-Britain** . . . . . 225  
 Christine Jacob, Laurence Maillard-Teyssier, Jean-Baptiste Denis and Caroline Bidot

16.1 Introduction . . . . . 225

16.2 Initial branching model . . . . . 226

16.3 Limit process as  $N_0 \rightarrow \infty$  . . . . . 229

16.4 Behavior of the BGW limit process . . . . . 233

    16.4.1 Extinction probability . . . . . 234

    16.4.2 Extinction time distribution . . . . . 235

    16.4.3 Size of the epidemic . . . . . 235

16.5 Estimation . . . . . 235

    16.5.1 Observations . . . . . 236

    16.5.2 Model and parameters . . . . . 236

    16.5.3 Prior distributions . . . . . 237

    16.5.4 Algorithm and software . . . . . 237

    16.5.5 Main results . . . . . 238

16.6 Conclusion . . . . . 239

References . . . . . 240

**17 Time to extinction of infectious diseases through age-dependent branching models** . . . . . 241  
 Miguel González, Rodrigo Martínez and Maroussia Slavtchova-Bojkova

17.1 Introduction . . . . . 241

17.2 Model of epidemic spread . . . . . 243

17.3 The epidemic’s time to extinction . . . . . 244

17.4 Determining vaccination policies . . . . . 246

    17.4.1 Vaccination based on the mean value of the time to extinction . . . . . 247

    17.4.2 Analyzing the control measures for avian influenza in Vietnam . . . . . 248

17.5 Concluding remarks . . . . . 251

17.6 Proofs ..... 251  
References ..... 256

**18 Time to extinction in a two-host interaction model for the macroparasite *Echinococcus granulosus* ..... 257**

Dominik Heinzmann  
18.1 Introduction ..... 257  
18.2 Prevalence-based interaction model ..... 258  
18.3 Approximating branching processes ..... 259  
18.4 Coupling ..... 260  
18.5 Time to extinction ..... 262  
18.6 Numerical illustration ..... 263  
References ..... 265

**Part VI Two-Sex Branching Models**

**19 Bisexual branching processes with immigration depending on the number of females and males ..... 269**

Shixia Ma and Yongsheng Xing  
19.1 Introduction ..... 269  
19.2 The bisexual process with immigration ..... 270  
19.3 The asymptotic growth rate ..... 271  
19.4 Limit behavior for the supercritical case ..... 273  
References ..... 276

**20 Two-sex branching process literature ..... 279**

Manuel Molina  
20.1 Introduction ..... 279  
20.2 The Daley’s two-sex branching process ..... 280  
20.3 Discrete time two-sex branching processes ..... 284  
    20.3.1 Processes with immigration ..... 284  
    20.3.2 Processes in varying or in random environments ..... 285  
    20.3.3 Processes depending on the number of couples in the population ..... 285  
    20.3.4 Processes with control on the number of progenitor couples ..... 287  
    20.3.5 Others classes of two-sex processes ..... 287  
20.4 Continuous time two-sex branching processes ..... 287  
20.5 Applications ..... 288  
    20.5.1 Application in the field of the Epidemiology ..... 289  
    20.5.2 Applications in the field of the Genetics ..... 289  
    20.5.3 Applications in population dynamics ..... 290  
20.6 Some suggestions for research ..... 290  
References ..... 291

**Index ..... 295**



# Contributors

Gerold Alsmeyer

Department of Mathematics and Computer Science, Institute of Mathematical Statistics, University of Münster, Einsteinstrasse 62, D-48149 Münster, Germany, e-mail: gerolda@math.uni-muenster.de

Frank Ball

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK, e-mail: frank.ball@nottingham.ac.uk

Caroline Bidot

UR 341, Applied Mathematics and Informatics unit, National Agronomical Research Institute, Jouy-en-Josas, France

Carlos A. Braumann

Department of Mathematics, Universidade de Évora, Rua Romão Ramalho 59, 7000-671 Évora, Portugal, e-mail: braumann@uevora.pt

Jean-Baptiste Denis

UR 341, Applied Mathematics and Informatics unit, National Agronomical Research Institute, Jouy-en-Josas, France

Miguel González

Department of Mathematics, University of Extremadura, Avda. Elvas s/n. 06006 Badajoz, Spain, e-mail: mvelasco@unex.es

Cristina Gutiérrez

Department of Mathematics, University of Extremadura, Avda. Elvas s/n. 06006 Badajoz, Spain, e-mail: cgutierrez@unex.es

Dominik Heinzmann

Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, e-mail: dominik.heinzmann@math.uzh.ch

Márton Ispány

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

Christine Jacob

UR 341, Applied Mathematics and Informatics unit, National Agronomical  
Research Institute, Jouy-en-Josas, France, e-mail: [christine.jacob@jouy.inra.fr](mailto:christine.jacob@jouy.inra.fr)

Marek Kimmel

Department of Statistics, Rice University, MS 138, 6100 Main Street, Houston,  
TX 77005, USA, and Systems Engineering Group, Institute of Automatic Control,  
Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland.  
e-mail: [kimmel@stat.rice.edu](mailto:kimmel@stat.rice.edu)

Fima C. Klebaner

School of Mathematical Sciences, Monash University, Clayton, Victoria 3800,  
Australia, e-mail: [fima.klebaner@sci.monash.edu.au](mailto:fima.klebaner@sci.monash.edu.au)

Shixia Ma

School of Science, Hebei University of Technology, 300401, Tianjin, China, e-mail:  
[shx.ma@eyou.com](mailto:shx.ma@eyou.com)

Laurence Maillard-Teyssier

UR 341, Applied Mathematics and Informatics unit, National Agronomical  
Research Institute, Jouy-en-Josas, France

Rodrigo Martínez

Department of Mathematics, University of Extremadura, Avda. Elvas s/n. 06006  
Badajoz, Spain, e-mail: [rmartinez@unex.es](mailto:rmartinez@unex.es)

Matthias Mathaes

Department of Statistics, Rice University, MS 138, 6100 Main Street, Houston, TX  
77005, USA

Penka Mayster

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences,  
Acad.G.Bonchev street, Bloc 8, 1113 Sofia, Bulgaria,  
e-mail: [penka.mayster@isetr.rnu.tn](mailto:penka.mayster@isetr.rnu.tn)

Kosto V. Mitov

Faculty of Aviation, 5856 D. Mitropolia, Pleven, Bulgaria, e-mail: [kmitov@af-acad.bg](mailto:kmitov@af-acad.bg)

Georgi K. Mitov

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 8 G.  
Bonchev street, 1113 Sofia, Bulgaria, e-mail: [gkmiov@gmail.com](mailto:gkmiov@gmail.com)

Manuel Molina

Department of Mathematics, Faculty of Sciences, University of Extremadura,  
06006 Badajoz, Spain, e-mail: [mmolina@unex.es](mailto:mmolina@unex.es)

Peter Neal

School of Mathematics, University of Manchester, Alan Turing Building, Oxford Road, Manchester M13 9PL, UK, e-mail: peter.neal@manchester.ac.uk

Gyula Pap

Faculty of Informatics, University of Debrecen, Pf.12, Debrecen H-4010 Hungary, e-mail: pap.gyula@inf.unideb.hu

Inés M. del Puerto

Department of Mathematics, University of Extremadura, Avda. Elvas s/n, 06006 Badajoz, Spain, e-mail: idelpuerto@unex.es

Ibrahim Rahimov

Department of Mathematics and Statistics, Zayed University, Dubai Campus, P.O.Box 19282, Dubai, UAE, e-mail: Ibrahim.Rahimov@zu.ac.ae

Maroussia Slavtchova-Bojkova

Department of Probability and Statistics. Institute of Mathematics and Informatics. Bulgarian Academy of Sciences, 1113-Sofia, Bulgaria, e-mail: bojkova@fmi.uni-sofia.bg

Valentin Topchii

Omsk Branch of Institute of Mathematics SB RAS, 13 Pevtsova street, Omsk 644099 Russia, e-mail: topchij@ofim.oscsbras.ru

Vladimir Vatutin

Steklov Mathematical Institute, Gubkin street 8, 119991 Moscow, Russia, e-mail: vatutin@mi.ras.ru

Yongsheng Xing

College of Mathematics, Shandong Institute of Technology, Yantai 264005, China, e-mail: xingys@nankai.edu.cn

George P. Yanev

Department of Mathematics, University of Texas - Pan American, 1201 West University Drive, Edinburg, Tx 78539, USA, e-mail: yanevgp@utpa.edu

Nikolay M. Yanev

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 8 G. Bonchev street, 1113 Sofia, Bulgaria, e-mail: yanev@math.bas.bg

## 6

# Extension of the Problem of Extinction on Galton-Watson Family Trees

George P. Yanev

**Abstract** We review the existing and present new results on certain subtrees of the Galton-Watson family tree. For a positive integer  $N$ , define an  $N$ -ary subtree to be the tree of a deterministic  $N$ -splitting, rooted at the ancestor. Dekking [2] raised and answered the question how to compute the probability for a branching process to possess the binary splitting property, i.e.,  $N = 2$ . Pakes and Dekking [8] studied the general situation when  $N \geq 2$ . Surprisingly, the case  $N \geq 2$  is studied so late, whereas the question for extinction of a branching process, i.e., non-existence of an infinite unary subtree ( $N = 1$ ) has been studied extensively over the past 120-150 years.

**Mathematics Subject Classification (2000):** 60J80, 05C05

**Keywords:** branching processes, Galton-Watson family trees, binary and  $N$ -ary trees, geometric offspring, Poisson offspring.

### 6.1 Introduction

Let  $\{Z_n\}$  be a Bienaymé-Galton-Watson process with offspring probability generating function (pgf)  $f(s) = \sum_{k=0}^{\infty} p_k s^k$ . Suppose  $Z_0 = 1$  and as usual assume  $p_k < 1$  for all  $k$ , and also that  $p_k > 0$  for some  $k > N$  where  $N$  is an integer. Galton [3] formulated his famous problem of “the decay of the families of men”, which is to determine the probability that the process  $\{Z_n\}$  becomes extinct, i.e.,  $Z_n = 0$  for some  $n \geq 1$ . Let us consider the family tree of  $\{Z_n\}$  (see [4], p.122-125 for a formal description). Note that the family tree is finite if and only if the process becomes extinct. As it was pointed out by Dekking in [2], the problem of non-extinction can be formulated as “with what probability does the family tree contain the (infinite)

---

George P. Yanev  
Department of Mathematics, University of Texas - Pan American, 1201 West University Drive,  
Edinburg, Texas 78539, USA, e-mail: yanevgp@utpa.edu

unary tree as a subtree, rooted at the root of the family tree?” Then he asked the question, “with what probability does the family tree of a branching process contain the (infinite) binary tree as a subtree, rooted at the root of the family tree (binary splitting property)?” Here the binary tree is the family tree associated to the branching process with offspring distribution given by  $p_2 = 1$ . It is surprising that this extension of the classical question for extinction of a branching process is studied so late, whereas the latter has been studied extensively over the past 120-150 years.

In this paper we study characteristics of certain subtrees of the family tree of  $\{Z_n\}$ . We call two subtrees disjoint if they do not have a common node different from the root of the entire tree. In addition, we consider only rooted and complete subtrees, i.e., subtrees rooted at the ancestor and being family trees of the deterministic branching process with pgf  $f(s) = s^N$ . Let the random variable  $V_N$  be the number of complete infinite and disjoint  $N$ -ary subtrees of a branching tree, rooted at the ancestor. The event  $\{V_1 > 0\}$  implies that there is at least one infinite unary subtree and thus the process would never die. The event  $\{V_2 > 0\}$  can be interpreted as the set of process’ trajectories where the family tree grows faster than binary splitting. Dekking [2] computes the probability for a branching process to possess the “binary splitting property”, i.e.,  $P(V_2 > 0)$ . Pakes and Dekking [8] study the general situation when  $N \geq 2$ . Mutafchiev [7] proves limit results for the survival probability of an  $N$ -ary subtree. In [11], Yanev and Mutafchiev study the distribution of  $V_N$ .

In Sections 2 and 3 we review some existing results. In Section 4 we prove a limit theorem for the ratio of the conditional expectations of  $Z_n$ , provided that an infinite  $N$ -ary and an unary subtree exist. Sections 5-7 cover the particular cases of geometric, Poisson, and “one-or-many” offspring distributions. There we discuss corollaries of the general results as well as some numerical illustrations. Finally, in the concluding remarks, we point out some links between the problem of existence of an infinite complete  $N$ -ary subtree and other research results.

## 6.2 Critical phenomenon

Define  $T_2 - 1$  to be the maximum height of a complete binary subtree rooted at the ancestor. Note that  $T_2 = 0$  if  $Z_1 < 2$ . Also,  $T_1 - 1$  is the maximum height of a unary subtree rooted at the ancestor and thus  $T_1$  is the extinction time of  $\{Z_n\}$ . We start this section with a theorem about the probability  $\gamma_2 = P(T_2 < \infty)$  that there is no infinite complete binary subtree, i.e., the growth is slower than binary splitting. Notice that  $\gamma_1 = P(T_1 < \infty)$  is the probability that there is no infinite unary subtree, i.e., the extinction probability of the process.

The following result is fundamental. We present the original Dekking’s proof.

**Theorem 6.1 (Dekking (1991)).** *The probability  $\gamma_2$  is the smallest root in  $[0, 1]$  of the fixed point equation*

$$x = g_2(x), \tag{6.1}$$

where

$$g_2(x) = f(x) + (1-x)f'(x).$$

*Proof.* For  $n = 1, 2, \dots$  let  $\gamma_2(n)$  be the probability that the family tree associated to  $\{Z_n\}$  does not contain a complete binary subtree of height  $n$  rooted at the ancestor. If the family tree has  $k$  nodes at level 1 (i.e.,  $Z_1 = k$ ), then it does not contain a complete binary subtree of height  $n+1$  iff  $k = 0$  or 1, or either all or all but one of the  $k$  subtrees rooted at these nodes do not contain a complete binary subtree of height  $n$ . Therefore one has

$$\begin{aligned} \gamma_2(n+1) &= p_0 + p_1 + \sum_{k=2}^{\infty} [\gamma_2^k(n) + k\gamma_2^{k-1}(n)(1-\gamma_2(n))]p_k \\ &= p_0 + p_1 + f(\gamma_2(n)) - p_0 - p_1\gamma_2(n) + (1-\gamma_2(n))(f'(\gamma_2(n)) - p_1), \end{aligned}$$

and hence

$$\begin{aligned} \gamma_2(n+1) &= f(\gamma_2(n)) + (1-\gamma_2(n))f'(\gamma_2(n)) \\ &= g_2(\gamma_2(n)). \end{aligned} \quad (6.2)$$

Therefore,  $\gamma_2 = \lim_{n \rightarrow \infty} \gamma_2(n)$  satisfies (6.1). Furthermore, if we put  $\gamma_0 = 0$ , then (6.2) is also true for  $n = 0$ . Since  $g_2'(x) = (1-x)f''(x)$  is nonnegative,  $g_2(x)$  is increasing and it follows that  $\gamma_2 = \lim_{n \rightarrow \infty} g_2^n(0)$  is the smallest root in  $[0, 1]$  of (6.1).  $\square$

Note that a similar recurrent argument can be applied to derive the equation  $x = f(x)$  for the probability of extinction  $\gamma_1$  (see Kemeny and Snell [6], p.80-81).

It turns out that no simple expression involving the moments of the offspring distribution can be found that will yield whether  $\gamma_2 = 1$  or not. In fact, Dekking [2] shows (see Theorem 6.2) by a counterexample that  $\gamma_2$  cannot be a continuous function of the moments of the offspring distribution, nor of any other parameter of the process which depends continuously on  $p_k$ ,  $k = 0, 1, 2, \dots$

In [8], Theorem 6.1 is generalized to the case of the presence of a complete  $N$ -ary subtree ( $N \geq 2$ ) rooted at the ancestor, i.e., the family tree of the deterministic branching process with pgf  $f(s) = s^N$ . The following theorem for the probability  $\gamma_N$  that there is no such  $N$ -ary subtree with infinite height holds.

**Theorem 6.2 (Pakes and Dekking (1991)).** *The probability  $\gamma_N$  is the smallest root in  $[0, 1]$  of the fixed point equation*

$$x = g_N(x), \quad (6.3)$$

where

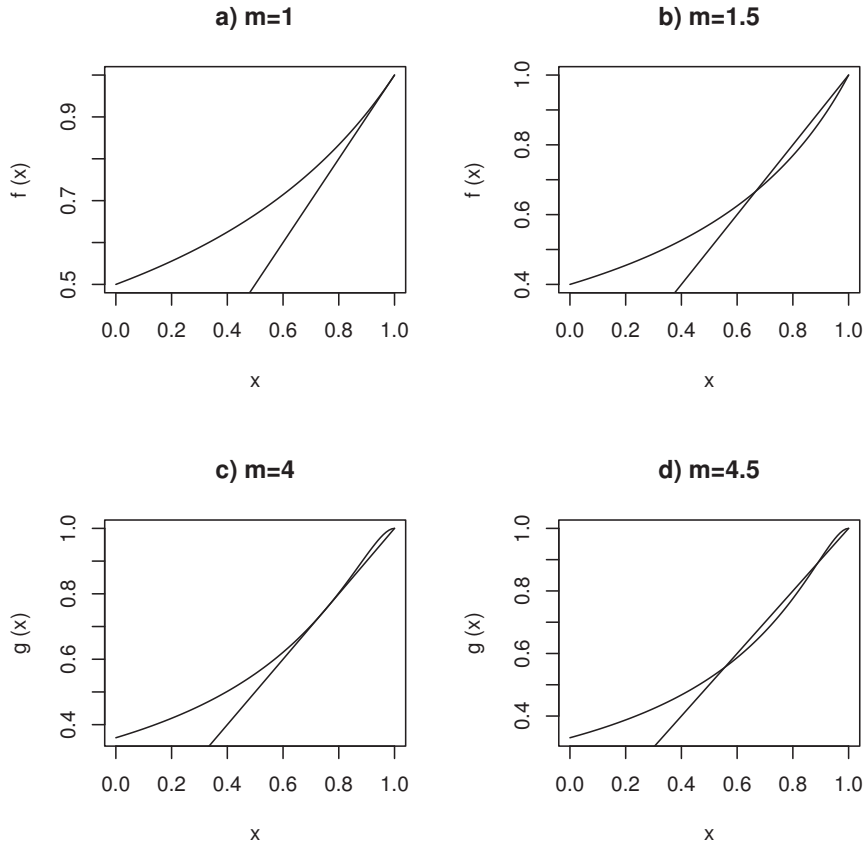
$$g_N(x) = \sum_{i=0}^{N-1} \frac{(1-x)^i}{i!} f^{(i)}(x).$$

In examples when the offspring distribution can be parameterized by its mean  $m$ , Pakes and Dekking [8] observe the following unusual critical phenomenon. There is an (critical) offspring mean value  $m_N^c > 1$  for  $N \geq 2$  such that

$$\begin{cases} \gamma_N = 1, m < m_N^c; \\ \gamma_N < 1, m \geq m_N^c. \end{cases}$$

This is qualitatively different to the behavior of the extinction (or not having an infinite unary subtree) probability  $\gamma_1$ , when the critical offspring mean is  $m_1^c = 1$  and

$$\begin{cases} \gamma_1 = 1, m \leq 1; \\ \gamma_1 < 1, m > 1. \end{cases}$$



**Fig. 6.1** Geometric offspring.

This difference occurs because  $g_N(x)$  is increasing but not convex;  $g_N'(1) = 0$ . Thus the critical case occurs when  $g_N(x)$  first touches the diagonal of the unit square somewhere other than unity. Figure 1 illustrates the situation for geometric offspring: a)  $m = 1$  - critical case for unary subtree; b)  $m = 1.5$  - supercritical case for unary subtree; c)  $m = 4$  - critical case for binary subtree; d)  $m = 4.5$  - supercritical case for binary subtree.

Next theorem gives a sufficient and a necessary condition for  $\gamma_N < 1$ .

**Theorem 6.3 (Pakes and Dekking (1991)).** (i) If  $\gamma_N < 1$ , then for some  $s \in (0, 1)$

$$(1-s)^{N-1} f^{(N)}(s) \geq (N-1)!$$

(ii) If

$$2N \sum_{k \geq N} \frac{p_k}{k+1} \leq \left( \sum_{k \geq N} p_k \right)^2,$$

then  $\gamma_N < 1$ .

We will finish this section with a remark from [8]. “Inspection of (6.3) should make it clear that there is virtually no hope of finding explicit expressions for the distribution of  $T_N$  for  $N \geq 2$ . It is even more unaccessible than the extinction time ( $T_1$ ) distribution.”

### 6.3 Distribution of the number of complete and disjoint subtrees, rooted at the ancestor

Recall that the random variable  $V_N$  for  $N = 1, 2, \dots$  is the number of disjoint complete  $N$ -ary subtrees with infinite height, rooted at the ancestor of  $\{Z_n\}$ . As the following result shows, the probability mass function of  $V_N$  can be obtained using the Taylor expansion of the pgf  $f(1)$  about the point  $\gamma_N$  evaluated at  $s = 1$ .

**Theorem 6.4 (Yanev and Mutafchiev (2006)).** For  $j = 0, 1, 2, \dots$  and  $N \geq 1$

$$P(V_N = j) = \sum_{k=jN}^{jN+N+1} \frac{(1-\gamma_N)^k}{k!} f^{(k)}(\gamma_N).$$

It is worth mentioning that  $P(V_N = j)$  is the  $(j+1)$ st segment of length  $N$  in the Taylor expansion of  $f(1)$  about  $\gamma_N$ . Indeed, set (for simplicity)  $N = 2$  and consider the Taylor expansion of  $f(1)$  about the point  $\gamma_2$ . Then

$$P(V_2 = 0) = f(\gamma_2) + (1-\gamma_2)f'(\gamma_2)$$

$$P(V_2 = 1) = \frac{(1-\gamma_2)^2}{2!} f''(\gamma_2) + \frac{(1-\gamma_2)^3}{3!} f'''(\gamma_2)$$

...

$$P(V_2 = j) = \frac{(1-\gamma_2)^{2j}}{(2j)!} f^{(2j)}(\gamma_2) + \frac{(1-\gamma_2)^{2j+1}}{(2j+1)!} f^{(2j+1)}(\gamma_2)$$

Denote by  $V_N(n)$  the number of complete, disjoint, and rooted subtrees of maximum height  $n$ , ( $n = 0, 1, 2, \dots$ ), i.e., for which  $T_N = n$ . Mutafchiev (2008) studies the survival probability  $P(V_N(n) > 0 | V_N = 0)$ . Observing that  $P(V_N(n) > 0) = P(T_N > n)$  and  $P(V_N(n) = 0) = P(T_N < \infty)$ , we can state his results as follows. Denote



$a_N = g'_N(\gamma_N)$  and  $2b_N = g''_N(\gamma_N)$ . One can see that

$$a_N = \frac{(1 - \gamma_N)^{N-1}}{(N-1)!} f^{(N)}(\gamma_N) \quad (6.4)$$

and if  $a_N = 1$ , then

$$2b_N = \frac{f^{(N+1)}(\gamma_N)}{f^{(N)}(\gamma_N)} - \frac{N-1}{1-\gamma_N}. \quad (6.5)$$

**Theorem 6.5 (Mutafchiev (2008)).** Assume  $\gamma_N \in (0, 1)$  for  $N \geq 2$ . Then  $a_N \leq 1$ .

(i) If  $a_N < 1$  (supercritical case), then as  $n \rightarrow \infty$

$$P(T_N > n \mid T_N < \infty) = c_N a_N^n + O(a_N^{2n}), \quad (6.6)$$

where  $c_N > 0$  is certain constant.

(ii) If  $a_N = 1$  (critical case) and  $b_N < \infty$ , then  $b_N > 0$  and as  $n \rightarrow \infty$

$$P(T_N > n \mid T_N < \infty) \sim \frac{1}{\gamma_N b_N n}. \quad (6.7)$$

*Remark 6.1.* (i) Note that if  $a_N = 1$  and  $f^{(N+1)}(1-) < \infty$ , then  $b_N < \infty$ , see (6.5). (ii) (6.6) extends to  $N \geq 2$  the classical results (e.g., Harris (1963), Theorem 8.4) for  $P(Z_n > 0) = P(T_1 > n)$  when  $m > 1$ . (iii) For  $P(Z_n > 0) = P(T_1 > n)$ , when  $m = 1$ , we have  $\gamma_1 = 1$  and hence (6.7) for  $N = 1$  is consistent with the Kolmogorov's result  $P(Z_n > 0) \sim 1/(b_1 n)$  as  $n \rightarrow \infty$ .

The subcritical case  $\gamma_N = 1$  ( $N \geq 2$ ) is treated in the following theorem.

**Theorem 6.6 (Pakes and Dekking (1991)).** Assume  $\gamma_N = 1$  for  $N \geq 2$  and that  $f^{(N)}(1-) < \infty$ . Then

$$P(T_N > n) \sim \exp\{-k_N N^n\} \quad (n \rightarrow \infty),$$

where  $k_N > 0$  is certain constant.

In Sections 5-7 we will come back to Theorem 6.5 discussing its corollaries for particular offspring distributions.

## 6.4 Ratio of expected values of $Z_n$ s provided infinite subtrees exist

In this section we prove a limit theorem for the ratio of the expected values of the  $n$ th generation's size,  $Z_n$ , as  $n \rightarrow \infty$ , conditioned on the existence of an  $N$ -ary and a unary subtree. Recall that  $a_N = g'_N(\gamma_N)$  and  $m = f'(1)$  is the offspring mean.

**Theorem 6.7.** Assume  $\gamma_N < 1$  for  $N \geq 1$ . If  $E(Z_1 \log(1 + Z_1)) < \infty$ , then

$$\lim_{n \rightarrow \infty} \frac{E(Z_n | T_N > n)}{E(Z_n | T_1 > n)} = \alpha_N \frac{1 - \gamma_1}{1 - \gamma_N} \quad (6.8)$$

where

$$\alpha_N = \frac{1}{m - a_N} \left[ m - \sum_{j=0}^{N-1} \frac{1}{j!} (1 - \gamma_N)^j f^{(j+1)}(\gamma_N) \right] \in (0, 1]. \quad (6.9)$$

*Proof.* Under the assumption  $E(Z_1 \log(1 + Z_1)) < \infty$ , there exists a random variable  $W$  such that  $Z_n/m^n \rightarrow W$  in  $L_1$  and  $EW = 1$  (e.g., [5], Theorem 2.7.3). Thus,

$$\alpha_N := \lim_{n \rightarrow \infty} E \left( \frac{Z_n}{m_n}; T_N > n \right) = E(W; T_N = \infty)$$

and, since  $\gamma_N < 1$ , we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{E(Z_n | T_N > n)}{E(Z_n | T_1 > n)} &= \lim_{n \rightarrow \infty} \frac{E(Z_n; T_N > n)P(T_1 > n)}{E(Z_n; T_1 > n)P(T_N > n)} \\ &= \lim_{n \rightarrow \infty} \frac{E(Z_n; T_N > n)P(T_1 > n)}{m^n P(T_N > n)} \\ &= \alpha_N \frac{1 - \gamma_1}{1 - \gamma_N}, \end{aligned}$$

which is (6.8). Next, we shall derive formula (6.9) for  $\alpha_N$ . Denote

$$\psi_n(s) = E(s^{Z_n}; T_N > n) \quad \text{and} \quad \phi_n(s) = E(s^{Z_n}; T_N \leq n).$$

From [8], Theorem 6.1, we have

$$\psi_{n+1}(s) = f_{n+1}(s) - \sum_{j=0}^{N-1} \frac{1}{j!} \psi_n^j(s) f^{(j)}(\phi_n(s)).$$

Differentiating with respect to  $s$ , we obtain

$$\psi'_{n+1}(s) = f'_{n+1}(s) - \sum_{j=1}^{N-1} \frac{1}{(j-1)!} \psi_n^{j-1}(s) \psi'_n(s) f^{(j)}(\phi_n(s)) - \sum_{j=0}^{N-1} \frac{1}{j!} \psi_n^j(s) f^{(j+1)}(\phi_n(s)) \phi'_n(s)$$

Setting  $s = 1$ , gives us

$$\begin{aligned} E(Z_{n+1}; T_N > n+1) &= EZ_{n+1} - E(Z_n; T_N > n) \sum_{j=1}^{N-1} \frac{1}{(j-1)!} (1 - \gamma_N)^{j-1} f^{(j)}(\gamma_N) \\ &\quad - [EZ_n - E(Z_n; T_N > n)] \sum_{j=0}^{N-1} \frac{1}{j!} (1 - \gamma_N)^j f^{(j+1)}(\gamma_N) \end{aligned}$$

Therefore,

$$\begin{aligned} E(Z_{n+1}; T_N > n+1) - E(Z_n; T_N > n) &= \frac{1}{(N-1)!} (1-\gamma_N)^{N-1} f^{(N)}(\gamma_N) \\ &= m^{n+1} - m^n \sum_{j=0}^{N-1} \frac{1}{j!} (1-\gamma_N)^j f^{(j+1)}(\gamma_N) \end{aligned}$$

Referring to (6.4), we have

$$E(Z_{n+1}; T_N > n+1) - a_N E(Z_n; T_N > n) = m^{n+1} - m^n \sum_{j=0}^{N-1} \frac{1}{j!} (1-\gamma_N)^j f^{(j+1)}(\gamma_N)$$

Dividing both sides by  $m^n$ , we find

$$\frac{m^{n+1}}{m^n} E\left(\frac{Z_{n+1}}{m^{n+1}}; T_N > n+1\right) - a_N E\left(\frac{Z_n}{m^n}; T_N > n\right) = m - \sum_{j=0}^{N-1} \frac{1}{j!} (1-\gamma_N)^j f^{(j+1)}(\gamma_N)$$

Passing to the limit as  $n \rightarrow \infty$ , we obtain

$$\alpha_N(m - a_N) = m - \sum_{j=0}^{N-1} \frac{1}{j!} (1-\gamma_N)^j f^{(j+1)}(\gamma_N) \quad (6.10)$$

Note that since  $\gamma_1 < \gamma_N < 1$ , we have  $m > 1$  and on the other hand by Theorem 6.5,  $a_N \leq 1$ . Hence  $m - a_N > 0$ . Dividing (6.10) by  $m - a_N$ , we arrive at (6.9).  $\square$

In the next three sections, we will prove corollaries of Theorem 6.7 for three particular cases of offspring distributions.

## 6.5 Geometric offspring distribution

In this section we present results for the case when the offspring distribution is geometric, i.e.,  $p_k = (1-p)p^k$ ,  $k \geq 0$  and  $p \in (0, 1)$ . Then for  $|s| < 1$  and  $N \geq 2$

$$f(s) = \frac{1-p}{1-ps} \quad \text{and} \quad g_N(s) = 1 - \left[ \frac{p(1-p)}{1-ps} \right]^N.$$

The number of  $N$ -ary subtrees  $V_N$  is geometric too (see [11]) given by

$$P(V_N = j) = \gamma_N (1-\gamma_N)^j, \quad j \geq 0,$$

where  $\gamma_N$  is the smallest solution in  $[0, 1]$  of

$$(1-x+m^{-1})^N = (1-x)^{N-1}. \quad (6.11)$$

It is clear from (6.11) that in the geometric case  $\gamma_N \rightarrow 0$  as  $m \rightarrow \infty$  and, as it is pointed out in [8],

$$\gamma_N = \frac{N}{m} + O(m^{-2}), \quad m \rightarrow \infty.$$

The critical value  $m_N^c$  for the offspring mean  $m$  is (see [8])

$$m_N^c = (N-1) \left(1 - \frac{1}{N}\right)^{-1} \quad (6.12)$$

and the corresponding probability of not having a  $N$ -ary subtree

$$\gamma_N^c = 1 - \left(1 - \frac{1}{N}\right)^N. \quad (6.13)$$

Note that if  $m = m_N^c$ , then as  $N \rightarrow \infty$

$$\gamma_N^c \rightarrow 1 - \frac{1}{e} \approx 0.6321 \quad \text{and} \quad EV_N^c = \frac{1 - \gamma_N^c}{\gamma_N^c} \rightarrow \frac{1}{e-1} \approx 0.5820.$$

In Table 1 we list some values of  $m_N^c$ ,  $\gamma_N^c$ , and  $EV_N^c$ .

$N$	2	3	4	6	10	20	100
$m_N^c$	4	6.750	9.481	14.930	25.812	53.001	270.468
$\gamma_N^c$	0.750	0.704	0.684	0.665	0.651	0.641	0.634
$EV_N^c$	0.333	0.421	0.463	0.504	0.535	0.559	0.577

**Table 6.1** Values of  $m_N^c$ ,  $\gamma_N^c$ , and  $EV_N^c$  for geometric offspring.

Next corollary of Theorem 6.5(ii) (critical case) is straightforward.

**Corollary 6.1.** *Assume geometric offspring. If  $m = m_N^c$ , where  $m_N^c$  is given by (6.12), then as  $n \rightarrow \infty$*

$$P(T_N^c > n \mid T_N^c < \infty) \sim \frac{2(1 - N^{-1})^{N-1}}{1 - (1 - N^{-1})^N} \frac{1}{n}.$$

Next result is a corollary of Theorem 6.7 for geometric offspring.

**Corollary 6.2.** *Assume geometric offspring. Then for any  $N = 2, 3, \dots$*

$$\lim_{n \rightarrow \infty} \frac{E(Z_n \mid T_N^c > n)}{E(Z_n \mid T_1^c > n)} = \frac{m-1}{m-a_N} (1 + a_N). \quad (6.14)$$

*Remark 6.2.* It is interesting to note that if  $a_N = 1$  (critical case) then the limit (6.14) equals the constant 2 for any  $N$ . This is rather unexpected.

*Proof.* We shall calculate  $\alpha(1 - \gamma_N)/(1 - \gamma_N)$ , which is the limit in Theorem 7. First, using (6.11) and  $p = n/(m + 1)$ , we obtain

$$\begin{aligned} a_N &= g'_N(\gamma_N) \\ &= \frac{(1 - \gamma_N)^{N-1} N! p^N (1 - p)}{(N - 1)! (1 - p\gamma_N)^{N+1}} \\ &= \frac{N}{m(1 - \gamma_N) + 1}. \end{aligned} \quad (6.15)$$

Also, since for  $j = 1, 2, \dots$

$$f^{(j)}(s) = \frac{j! p^j (1 - p)}{(1 - ps)^{j+1}},$$

we have

$$\begin{aligned} \sum_{j=0}^{N-1} \frac{1}{j!} (1 - \gamma_N)^j f^{(j+1)}(\gamma_N) &= \sum_{j=0}^{N-1} \frac{1}{j!} (1 - \gamma_N)^j \frac{(j+1)! (1 - p) p^{j+1}}{(1 - p\gamma_N)^{j+2}} \\ &= \frac{p(1 - p)}{(1 - p\gamma_N)^2} \sum_{j=0}^{N-1} (j+1) \left[ \frac{p(1 - \gamma_N)}{1 - p\gamma_N} \right]^j \\ &= \frac{m}{(1 + m - m\gamma_N)^2} \sum_{j=0}^{N-1} (j+1) \left[ \frac{m(1 - \gamma_N)}{m + 1 - m\gamma_N} \right]^j \\ &= \frac{1}{(1 + m(1 - \gamma_N))^2} \left[ 1 + \sum_{j=0}^{N-1} (j+1) \left[ \frac{m(1 - \gamma_N)}{1 + m(1 - \gamma_N)} \right]^j \right]. \end{aligned}$$

Denote  $S_N(x) = \sum_{j=1}^{N-1} (j+1)x^j$  for some  $x > 0$ . One can see that

$$\begin{aligned} S_N(x) - xS_N(x) &= 2x + \sum_{i=2}^{N-1} x^i - Nx^N \\ &= 2x + x^2 \frac{1 - x^{N-2}}{1 - x} - Nx^N \end{aligned}$$

and thus,

$$\begin{aligned} 1 + S_N(x) &= 1 + \frac{2x}{1 - x} + x^2 \frac{1 - x^{N-2}}{(1 - x)^2} - \frac{Nx^N}{1 - x} \\ &= \frac{Nx^{N+1} - (N+1)x^N + 1}{(1 - x)^2}. \end{aligned}$$

Set  $x = m(1 - \gamma_N)/[1 + m(1 - \gamma_N)]$  and thus  $1 - x = [1 + m(1 - \gamma_N)]^{-1}$ . Now, after some algebra and using (6.11) and (6.15), we obtain

$$\begin{aligned}
\alpha_N(m - a_N) &= m - \sum_{j=0}^{N-1} \frac{1}{j!} (1 - \gamma_N)^j f^{(j+1)}(\gamma_N) \\
&= m[1 - (Nx^{N+1} - (N+1)x^N + 1)] \\
&= m \left[ \frac{m(1 - \gamma_N)}{1 + m(1 - \gamma_N)} \right]^N \left[ N + 1 - N \frac{m(1 - \gamma_N)}{1 + m(1 - \gamma_N)} \right] \\
&= \frac{m^{N+1}(1 - \gamma_N)^N}{[1 + m(1 - \gamma_N)]^N} \frac{N + 1 + m(1 - \gamma_N)}{1 + m(1 - \gamma_N)} \\
&= m(1 - \gamma_N) \frac{N + 1 + m(1 - \gamma_N)}{1 + m(1 - \gamma_N)}.
\end{aligned} \tag{6.16}$$

Therefore, using (6.15), (6.16), and  $\gamma_1 = 1/m$ , we have for the limit in Theorem 7

$$\begin{aligned}
\alpha_N \frac{1 - \gamma_1}{1 - \gamma_N} &= \frac{m(1 - \gamma_1)}{m - a_N} \left[ 1 + \frac{N}{1 + m(1 - \gamma_N)} \right] \\
&= \frac{m - 1}{m - a_N} (1 + a_N),
\end{aligned}$$

which completes the proof of the corollary.  $\square$

## 6.6 Poisson offspring distribution

Consider Poisson offspring distribution with pgf  $f(s) = e^{m(s-1)}$ ,  $m > 0$ . Whence

$$g_N(s) = e^{-m(1-s)} \sum_{j=0}^{N-1} \frac{1}{j!} [m(1-s)]^j. \tag{6.17}$$

The distribution of  $V_N$  (see [11]) can be presented for  $j = 0, 1, \dots$  as

$$P(V_N = j) = P(jN \leq Y_N \leq jN + N - 1),$$

where  $Y_N$  is a Poisson random variable with mean  $m(1 - \gamma_N)$ . That is,  $P(V_N = j)$  is the  $(j + 1)$ st segment of length  $N$  in the distribution of  $Y_N$ . Some values of  $EV_N^c$  are given in Table 2. In the Poisson case, one also has  $\gamma_N \rightarrow 0$  as  $m \rightarrow \infty$ . Now, (6.3) and (6.17) lead to

$$\gamma_N \sim \frac{m^{N-1} e^{-m}}{(N-1)!}, \quad m \rightarrow \infty.$$

Denote  $y = m_N^c (1 - \gamma_N^c)$ . It is shown in [8] that  $m_N^c$  and  $\gamma_N^c$  satisfy the equations

$$\frac{y^N}{(N-1)!} + \sum_{j=0}^{N-1} \frac{y^j}{j!} = e^y \tag{6.18}$$

and

$$m_N^c y^{N-1} = (N-1)! e^y. \quad (6.19)$$

Since (6.18) has a unique positive solution, one can find  $m_N^c$  and  $\gamma_N^c$  from (6.18) and then (6.19). Some values are given in Table 2 below. It is proved in [8] that as  $N \rightarrow \infty$

$$m_N^c = N + \sqrt{N \log N} (1 + o(1)) \quad \text{and} \quad \gamma_N^c = \sqrt{\pi/(2N)} (1 + o(1)).$$

$N$	2	3	4	6	10	20	40
$m_N^c$	3.351	5.150	6.800	9.876	15.582	28.775	53.434
$\gamma_N^c$	0.465	0.343	0.282	0.218	0.160	0.107	0.072
$EV_N^c$	0.654	0.793	0.844	0.876	0.886	0.899	0.929

**Table 6.2** Values of  $m_N^c$ ,  $\gamma_N^c$ , and  $EV_N^c$  for Poisson offspring.

Here we state a straightforward corollary of Theorem 6.5(ii) (critical case).

**Corollary 6.3.** *Assume Poisson offspring. If  $m = m_N^c$ , then as  $n \rightarrow \infty$*

$$P(T_N^c > n \mid T_N^c < \infty) \sim \frac{2(1 - \gamma_N^c)}{m_N^c(1 - \gamma_N^c) - N + 1} \frac{1}{n}.$$

Next result is a corollary of Theorem 6.7 for Poisson offspring.

**Corollary 6.4.** *Assume Poisson offspring. Then for any  $N = 2, 3, \dots$*

$$\lim_{n \rightarrow \infty} \frac{E(Z_n \mid T_N > n)}{E(Z_n \mid T_1 > n)} = \frac{m(1 - \gamma_1)}{m - a_N}. \quad (6.20)$$

*Remark 6.3.* . (i) The limit in (6.20) can be expressed in the form

$$(1 - \gamma_1) \left\{ 1 - \frac{[m(1 - \gamma_N)]^{N-1}}{(N-1)!} e^{-m(1 - \gamma_N)} \right\}^{-1}.$$

(ii) Let  $m_N = m_N^c$ . Then  $a_N^c = 1$  and since the probability of extinction  $\gamma_1^c$  satisfies  $\gamma_1^c = \exp\{-m_N^c(1 - \gamma_1^c)\}$ , one can see that for the limit in (6.20) it is true that

$$\frac{m_N^c(1 - \gamma_1^c)}{m_N^c - 1} > 1 \quad \text{and} \quad \frac{m_N^c(1 - \gamma_1^c)}{m_N^c - 1} \downarrow 1 \quad \text{as} \quad N \rightarrow \infty.$$

*Proof.* We have  $f(s) = \exp\{-m(1 - s)\}$  and thus for  $j = 1, 2, \dots$

$$f^{(j+1)}(s) = m^{j+1}e^{-m(1-s)} = mf^{(j)}(s)$$

Therefore, referring to (6.9) and (6.17),

$$\begin{aligned} \alpha_N &= \frac{1}{m-a_N} \left[ m - \sum_{j=0}^{N-1} \frac{1}{j!} (1-\gamma_N)^j f^{(j+1)}(\gamma_N) \right] \\ &= \frac{m}{m-a_N} \left[ 1 - m^{-1} \sum_{j=0}^{N-1} \frac{1}{j!} (1-\gamma_N)^j m f^{(j)}(\gamma_N) \right] \\ &= \frac{m}{m-a_N} \left[ 1 - e^{-m(1-\gamma_N)} \sum_{j=0}^{N-1} \frac{1}{j!} [m(1-\gamma_N)]^j \right] \\ &= \frac{m(1-\gamma_N)}{m-a_N}. \end{aligned}$$

Hence,

$$\alpha_N \frac{1-\gamma_1}{1-\gamma_N} = \frac{m(1-\gamma_1)}{m-a_N}$$

and the limit in (6.20) follows from Theorem 6.7.  $\square$

## 6.7 One-or-many offspring distribution

In this section we consider a two-parameter family of offspring distributions  $\{p_k\}$  defined for  $p \in (0, 1)$  and some integer  $r > N > 1$  by  $p_1 = 1-p$ ,  $p_k = 0$  for  $2 \leq k \leq r-1$ , and  $p_r = p$ . Clearly  $m = 1-p+rp$  and  $f(s) = (1-p)s + ps^r$ . Hence

$$g_N(s) = 1-p \sum_{j=N}^r \binom{r}{j} (1-s)^j s^{r-1}.$$

Consequently, Theorem 6.2 shows that  $\gamma_N$  is the smallest solution in  $[0, 1]$  of

$$s = p \sum_{j=N}^r \binom{r}{j} s^j (1-s)^{r-j}.$$

Let  $B_r(\gamma_N)$  denote a binomial  $(r, 1-\gamma_N)$  random variable. It is shown in [11] that  $P(V_N = 0) = 1-p+pP(B_r(\gamma_N) \leq N-1)$  and for  $j = 1, 2, \dots$

$$P(V_N = j) = \begin{cases} pP(jN \leq B_r(\gamma_N) \leq jN+U) & \text{if } jN \leq r, \\ 0 & \text{if } jN > r, \end{cases}$$

where  $U = \min\{N-1, r-jN\}$ . That is,  $P(V_N = j)$  is the  $(j+1)$ st segment of length  $U$  in the distribution of  $B_r(\gamma_N)$ . It is shown in [8] that  $\gamma_N^c$  is the unique solution of



$$\sum_{j=N}^r \binom{r}{j} \left(\frac{1-x}{x}\right)^{j-N} = r \binom{r-1}{N-1}$$

and also for the critical value  $p_N^c$

$$(p_N^c)^{-1} = r \binom{r-1}{N-1} (1-\gamma_N^c)^{N-1} (\gamma_N^c)^{r-N}. \quad (6.21)$$

In particular, if  $r = N + 1$  then

$$m_N^c = 1 + (N-1) \left(1 - \frac{1}{N^2}\right)^{-N}, \quad \gamma_N^c = N^{-2}, \quad \text{and} \quad EV_N^c = 1 - N^{-2}. \quad (6.22)$$

We list some values of  $m_N^c$ ,  $\gamma_N^c$ , and  $EV_N^c$  in Table 3.

$N$	2	3	4	6	10	20	40
$m_N^c$	2.778	3.848	4.884	6.921	10.952	20.975	40.988
$\gamma_N^c$	0.250	0.111	0.063	0.028	0.010	0.003	0.001
$EV_N^c$	0.750	0.889	0.938	0.972	0.990	0.998	0.999

**Table 6.3** Values of  $m_N^c$ ,  $\gamma_N^c$ , and  $EV_N^c$  for one-or- $(N+1)$  offspring.

The following straightforward corollary of Theorem 6.5(ii) holds.

**Corollary 6.5.** *Assume one-or- $(N+1)$  offspring distribution. If  $m = m_N^c$ , then*

$$P(T_N^c > n \mid T_N^c < \infty) \sim 2 \left(N - \frac{1}{N}\right) \frac{1}{n} \quad (n \rightarrow \infty).$$

Next result is a corollary of Theorem 6.7 for one-or- $(N+1)$  offspring.

**Corollary 6.6.** *Assume one-or- $(N+1)$  offspring distribution. If  $m = m_N^c$ , then*

$$\lim_{n \rightarrow \infty} \frac{E(Z_n \mid T_N^c > n)}{E(Z_n \mid T_1^c > n)} = \frac{N}{N-1} \left(1 - \frac{2^N - 1}{N^{2N}}\right).$$

*Proof.* We have for  $j \geq 1$

$$f'(s) = 1 - p + p(N+1)s^N \quad \text{and} \quad f^{(j+1)}(s) = p(N+1)N \dots (N+1-j)s^{N-j} \quad (6.23)$$

It follows from (6.9) and (6.21)-(6.23), after some algebra, that

$$\begin{aligned}
\alpha_N &= \frac{1}{m_N^c - 1} \left[ m_N^c - f'(\gamma_N^c) - \sum_{j=1}^{N-1} \frac{1}{j!} (1 - \gamma_N^c) f^{(j+1)}(\gamma_N^c) \right] \\
&= \frac{(1 - N^{-2})^N}{(N-1)} \left[ \frac{N - N^{-1}}{(1 - N^{-2})^N} - \frac{1 - N^{-1}}{(1 - N^{-2})^N N^{2N}} \sum_{j=0}^{N-1} \frac{(N+1) \dots (N+1-j)}{j!} \right] \\
&= (1 + N^{-1}) - \frac{1}{N^{2N+1}} \sum_{j=0}^{N-1} \frac{(N+1)N \dots (N+1-j)}{j!} \\
&= (1 + N^{-1}) - \frac{(N+1)(2^N - 1)}{N^{2N+1}}
\end{aligned}$$

and the corollary follows from Theorem 6.7.  $\square$

## 6.8 Concluding Remarks

Pakes and Dekking [8] point out that constructions of complete  $N$ -ary subtrees of a Galton-Watson family tree, have appeared in the study of Mandelbrot's percolation process by Chayes et. al. [1] and in Pemantle's work [9] on reinforced random walks. In particular, Pemantle's results imply that: if there exists  $s_0 \in (0, 1)$ , such that  $g_N(s_0) \leq s_0$ , then  $\gamma_N \leq s_0 < 1$ . In [7], Mutafchiev discusses an interesting connection between the subject matter  $N$ -ary trees and the existence of a giant  $k$ -core in a random graph. Consider an Erdős-Rényi random graph  $G(n, p)$  with  $n$  vertices in which the possible arcs are present independently, each with probability  $p$ . Pittel et. al. [10] construct a Galton-Watson family tree rooted at a vertex of the graph  $G(n, \lambda/n)$ , ( $\lambda > 0$ ), assuming Poisson offspring distribution with mean  $\lambda$ . They show that a giant  $k$ -core appears suddenly when the number of arcs reaches  $c_k n/2$ , where the constant  $c_k$  can be explicitly computed. It is remarkable that the values of  $c_k$  coincide with those of  $m_{k-1}^c$  for  $k = 3, 4, 5$  in case of Poisson offspring distribution (see Table 2 and [10], p.114). This needs further investigation.

## References

1. Chayes, J.L., Chayes, L., and R. Durrett, R.: Connectivity properties of Mandelbrot's percolation process. *Prob. Theor. Rel. Fields* **77**, 307-324, (1988)
2. Dekking, F.M.: Branching processes that grow faster than binary splitting. *Amer. Math. Monthly* **98**, 728-731 (1991)
3. Galton, F.: Problem 4001. *Educational Times* **17**, (1873)
4. Harris, T.E.: *The Theory of Branching Processes*, Springer, Berlin (1963)
5. Jagers, P.: *Branching Processes with Biological Applications*, Wiley, London (1975)
6. Kemeny, J.G. and Snell, J.L.: *Mathematical Models in the Social Sciences*, Reprint of the 1962 ed., issued in series: Introduction to higher mathematics, The MIT Press, Cambridge, MA. (1972)

7. Mutafchiev, L.R.: Survival probabilities for  $N$ -ary subtrees on a Galton-Watson family tree. *Statist. Probab. Lett.* **78**, 2165-2170 (2008)
8. Pakes, A.G., Dekking, F.M.: On family trees and subtrees of simple branching processes. *J. Theor. Probab.* **4**, 353-369 (1991)
9. Pemantle, R.: Phase transition in reinforced random walk and RWRE on trees. *Ann. Probab.* **16**, 1229-1241 (1988)
10. Pittel, B., Spencer, J. Wormald, N.: Sudden emergence of a giant  $k$ -core in a random graph. *J. Combin. Theory Ser. B* **67**, 111-151 (1996)
11. Yanev, G.P. and Mutafchiev, L.R.: Number of complete  $N$ -ary subtrees on Galton-Watson family trees. *Methodol. Comput. Appl. Probab.* **8**, 223-233 (2006)