

Haldane's Asymptotics for Slightly Supercritical Processes

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 12
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von

Florin Boenkost
aus Frankfurt am Main

Frankfurt am Main (2021)

(D 30)

vom Fachbereich 12 (Informatik und Mathematik) der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Martin Möller

Gutachter: Prof. Dr. Anton Wakolbinger
Dr. Cornelia Pokalyuk

Datum der Disputation: 16.12.2021

Acknowledgements

First and foremost I would like to express my deepest gratitude to both my advisors Anton Wakolbinger and Cornelia Pokalyuk, who guided me throughout my doctoral studies at Goethe-University. Anton, who with his enthusiasm sparked my interest in mathematical population genetics and gave me the opportunity to work with other accomplished scientist. Cornelia, with whom I had numerous stimulating discussion about mathematics and life itself.

Secondly, I would like to thank all my other collaborators from the different projects resulting in this thesis, Adrián González Casanova, Götz Kersting, Matthias Birkner and Iulia Dahmer, from whom I learnt so much about various topics.

I also would like to express my gratitude to the whole probability group in Frankfurt for so many interesting talks and events, I would like to mention here Ralph Neininger for introducing me to probability theory and for organizing the *Blockseminar Riezlern*, which was always a pleasure. I am much obliged to Anna Weiglhofer for her technical support in the last years.

I am tremendously thankful to my family and friends, especially to my parents Karin and Wolfram who supported me and always gave helpful advise throughout this years.

To all my colleagues and fellow students with whom I enjoyed so many lunch and coffee breaks, thank you Alexander Molitor, Christopher Lorenz, Elias Polak, Insea Schlattmeier, Jasmin Straub, Joel Kübler, Jonas Knörr, Jonas Peter, Leon Fröber, Oliver Gebhard, Raphael Eichhorn and Solveig Plomer.

Contents

1	Introduction	1
2	Framework and synopsis of results	5
2.1	Branching processes and population genetics	5
2.2	Haldane's asymptotics for Cannings models with moderate selection	12
2.2.1	A graphical representation including selection	12
2.2.2	The Cannings ancestral selection process	15
2.2.3	The case of moderately weak selection	17
2.2.4	The case of moderately strong selection	20
2.2.5	Moderate selection and heavy-tailed offspring distributions	23
2.3	Haldane's asymptotics for branching processes in an iid random environment	27
3	Haldane's formula in Cannings models:	
	The case of moderately weak selection	33
3.1	Introduction	33
3.2	Cannings models with selection	35
3.2.1	A paintbox representation for the neutral reproduction	35
3.2.2	A paintbox representation incorporating selection	35
3.2.3	The Cannings ancestral selection process	36
3.3	Main results	37
3.3.1	Duality of Cannings frequency and ancestral selection process	37
3.3.2	Haldane's formula for Cannings models with selection	37
3.4	Haldane's formula in the Moran model	40
3.5	The Cannings ancestral selection graph. Proof of Theorem 7	41
3.6	Coupling of the Cannings and Moran ancestral selection processes. Proof of Theorem 8a	45
3.7	A concentration result for the equilibrium distribution of the CASP. Proof of Theorem 8	58
4	Haldane's formula in Cannings models:	
	The case of moderately strong selection	71
4.1	Introduction	71
4.2	A class of Cannings models with selection	73
4.2.1	Paintbox representation in the neutral case	73
4.2.2	A paintbox representation with selection	73
4.2.3	The Cannings frequency process	74
4.3	Main Result	74
4.4	Auxiliary results	79
4.4.1	Slightly supercritical Galton-Watson processes	79
4.4.2	Estimates on the paintbox	82
4.5	Proof of the main result	85
4.5.1	First phase: From 1 to $N^{b+\delta}$	86

4.5.2	Second phase: from $N^{b+\delta}$ to εN	90
4.5.3	Third phase: from εN to N	94
4.6	Discussion	94
5	Haldane's asymptotics for Supercritical Branching Processes in an iid Random Environment	97
5.1	Introduction and main result	97
5.2	An expression for the survival probability	100
5.3	On perpetuities with small interest rates	102
5.4	Proof of Theorem 10	108
	Bibliography	117
A	Zusammenfassung	125
A.1	Haldanes Asymptotik für leicht superkritische Prozesse	125
A.1.1	Einleitung	125
A.1.2	Cannings-Modelle mit moderater Selektion	128
A.1.3	Der anzestrale Selektionsprozess im Cannings-Modell	131
A.1.4	Der Fall moderat schwacher Selektion	133
A.1.5	Der Fall moderat starker Selektion	134
A.1.6	Moderate Selektion und Nachkommenverteilungen mit schweren Rändern	134
A.1.7	Haldanes Asymptotik für Verzweigungsprozesse in zufälliger u.i.v. Umgebung	136

List of Symbols

$\mathbb{E}[X]$		Expectation of the random variable X .
$\text{Var}(X)$		Variance of the random variable X .
$\mathcal{L}(X)$		Distribution of the random variable X .
$X \stackrel{d}{=} Y$		X and Y are equal in distribution.
\mathbb{N}_0	$= \{0, 1, 2, \dots\}$	Set of all natural numbers including 0.
\mathbb{Z}	$= \{\dots, -2, -1, 0, 1, 2, \dots\}$	Set of all integer numbers.
\mathbb{R}	$= (-\infty, \infty)$	Set of all real numbers.
$[n]$	$= \{1, \dots, n\}$	Set of all natural numbers between 1 and n .
$\lfloor x \rfloor$	$= \max\{k \in \mathbb{Z} : k \leq x\}$	Largest integer part of x .
$\lceil x \rceil$	$= \min\{k \in \mathbb{Z} : k \geq x\}$	Smallest integer bigger or equal than x .

Chapter 1

Introduction

Mathematical population genetics is the synthesis of the Darwinian theory of evolution [Dar59] and Mendel's laws of inheritance [Men65] started in the early twentieth century by its founding fathers Fisher, Haldane and Wright.

The theory of population genetics aims to describe, among other topics, the evolution of a population (or gene) in an idealised mathematical framework, to answer questions arising both from a biological or from a mathematical background. For an introduction to mathematical population genetics the reader is referred to the monographs by Durrett [Dur08], Etheridge [Eth11] and Ewens [Ewe04].

There are many forces acting on populations and influencing the frequency of an allele or a specific type of individuals, for example genetic drift, selection, mutation and migration to name only a few. In this thesis we are mostly concerned with the interplay between selection and genetic drift and the resulting consequences for the probability of survival and eventual fixation of a single beneficial mutant with small selective advantage. Genetic drift refers to the random change in frequency of a gene due to the random reproduction of individuals and should not be confused with the notion of the drift of a diffusion process that is common in probability theory.

One of the widest known models in population genetics is the *Wright-Fisher* model [Fis30] and [Wri31], where the evolution of N haploid individuals is modelled in the following way. In each of the discrete non-overlapping generations every child chooses a parent independently and uniformly among the N possible individuals in the previous generation, which results in multinomially distributed vector of offspring numbers. Let us assume there are two types of individuals, *wildtype* ones and individuals of *beneficial type* and assume that children inherit the type of their parent. We incorporate directional selection with strength $s > 0$ into this model, by decreeing that children choose a parent proportional to their weight, where a wildtype individual has weight 1 and a beneficial individual has the increased weight $1 + s$. Therefore, the beneficial individuals are favoured compared to the wildtype individuals and if the beneficial type is rare, each beneficial individual has in expectation approximately $1 + s$ many offspring.

A natural question arising in these models is whether the beneficial type conquers the whole population and eventually fixates or dies out, once the beneficial type enters the population. This founding event might happen due to a migration or mutation, however in the following we are not concerned with the origin of the beneficial type and assume that no further beneficial individuals come into play by these mechanisms.

The study of survival probabilities of stochastic processes especially of Galton-Watson processes, has a long history and goes back to Galton and Watson [Gal73], [GW74] and the at first overlooked work by Bienaymé [Bie45]. They were interested in the probability of extinction of family names, see [Ken66] and [Ken75] for a short summary on the early history of Galton-Watson processes.

In a population genetic setting, Haldane [Hal27] argued that the fixation probability $\pi(s)$ of a single mutant gene in a large population with small selective advantage s obeys the asymptotics

$$\pi(s) \sim 2s \quad \text{as } s \downarrow 0.$$

For the early phase in which the selective mutant is sufficiently rare, Haldane approximated the binomially distributed offspring numbers in the Wright-Fisher model by a Galton-Watson process, where each individual has $\text{Pois}(1+s)$ many offspring. Since the variance of $\text{Pois}(1+s)$ converges to 1 as $s \downarrow 0$, the factor 2 in the above asymptotics equals $2/\sigma^2$. Thus, *Haldane's asymptotics* can be read as

$$\pi(s) \sim \frac{2s}{\sigma^2} \quad \text{as } s \downarrow 0,$$

see [PW08] for a historical overview. In the context of Galton-Watson processes this asymptotics was also observed by Kolmogorov [Kol38] and further investigated by Athreya [Ath92], Eshel [Esh81], [Esh84], Hoppe [Hop92] and others.

The Wright-Fisher model was generalised by Cannings in the 1970s, by allowing an exchangeable number of offspring, such that the population size stays constant [Can74] and [Can75]. In [BGPW21b], which is Chapter 4 in this thesis, selection is incorporated to a rather large subclass of Cannings models in a similar vein as in Wright-Fisher models with selection. This approach differs from the ones taken by other authors, see e.g. [LL07], but has the advantage of allowing a forward and backward representation of the Cannings models. We then prove Haldane's asymptotics for the regime of *moderately strong selection*. This is achieved in a manner similar to Haldane's original program, by comparing the Cannings model with selection to supercritical Galton-Watson processes. For the case of Cannings models with *moderately weak selection* a backwards approach turns out to be successful for proving Haldane's asymptotics. This has been accomplished in [BGPW21a], which is Chapter 3 in this thesis. This approach can be seen as a continuation of the historical developments described in the sequel.

At a similar time as Cannings, Moran [Mor58] introduced a population model in continuous time, which allows for overlapping generations, now known as the *Moran model*. For each pair of individuals reproduction events occur at rate 1. At reproduction with equal probability one individual among the two replaces the other one with its offspring, which inherits the type of its parent. Assume each individual has one of two types, the wildtype or the beneficial type. Additional to the just described *neutral reproduction events*, individuals experience *selective reproduction events* at rate s , at which a randomly picked individual is replaced by the offspring of the reproducing individual. However, these events are only realised when the individual is of beneficial type.

All these models introduced so far describe the evolution of a population forwards in time. If one is interested in genealogies and in ancestries of a population, one arrives at processes moving backwards in time. The most basic example for a backwards process is *Kingman's coalescent* introduced in the 1980s [Kin82a]. Kingman's coalescent is a partition valued process in continuous time, where each pair of blocks merges independently at rate 1. If one takes n individuals in the neutral Wright-Fisher model or in the neutral Moran model, then its genealogy, when properly rescaled, converges to Kingman's coalescent [Kin82b]. However, Kingman's coalescent describes the limiting genealogy only when dealing with neutral evolution.

Related to Kingman's coalescent is the *ancestral selection graph* introduced by Krone and Neuhauser [KN97], [NK97] for the Moran model with selection. Starting with n individuals each pair of lines of the ancestral selection graph coalesces whenever a neutral reproduction event occurs among the n individuals. In addition each ancestral lineage branches into two potential ancestral lineages, whenever one of the n individuals is targeted to be replaced by an individual outside of the sample at an selective reproduction event. Ancestral processes have been studied in a wide range of models, not exclusively of models involving selection. They have been extended to models incorporating mutation, recombination or others, see [PP13], [BLW16], [BB20], [Cor17] to only name a few.

Returning to the case of moderately weak selection in Cannings models, the scope of [BGPW21a], which is Chapter 3 in this thesis, is twofold. First, we introduce the *Cannings ancestral selection graph*, which describes the genealogy of a Cannings model with selection and thus is an analogue of the Krone-Neuhauser ancestral selection graph for Cannings models. We prove that the Cannings ancestral selection process, which is the block counting process of the Cannings ancestral selection graph, is in sampling duality to the *Cannings frequency process* (counting the number of wildtype individuals). Secondly, Haldane's asymptotics for Cannings models with moderate selection is proven by making use of the aforementioned duality.

As the last part of this thesis, in the footsteps of Haldane and Kolmogorov, we analyse the survival probability of supercritical branching processes in an iid random environment ([BK21] which is Chapter 5 in this thesis). Branching processes in a random environment are a natural generalisation of Galton-Watson processes and were first introduced around the 1970s by Smith and Wilkinson [SW69] and Athreya and Karlin [AK71b], [AK71a]. In contrast to classical Galton-Watson processes the offspring distribution is now allowed to change between the generations in an independent fashion, which results in randomly fluctuating offspring means between the generations. Under assumptions ensuring that the fluctuations of the random offspring mean are not too large, we prove Haldane's asymptotics for supercritical branching processes in an iid random environment and furthermore obtain a smooth adaption of Haldane's asymptotics for the survival probability as the branching process comes closer to criticality. In the proof we make use of an expression for the survival probability of branching processes in random environment, which does not rely on the classical fixed point characterization. This expression discloses similarities to perpetuities known from a financial context. We obtain a limiting theorem for perpetuities with vanishing interest rates. This result seems to be of interest in its own right.

The subsequent part of this thesis is organised as follows. Chapter 2 is divided into two parts. Section 2.1 introduces some mathematical background and further describes the scope of this thesis. Afterwards, starting with Section 2.2 we give a synopsis of the results and techniques used in [BGPW21b], [BGPW21a] and [BK21] and conclude with an outlook on so far unpublished work in progress joint with Matthias Birkner, Iulia Dahmer and Cornelia Pokalyuk in Section 2.2.5.

Chapters 3 to 5 are identical with the submitted or published versions of the corresponding papers, up to some small changes in the layout. The results of Chapter 3 and 4 have been obtained in collaboration with Adrián González Casanova, Cornelia Pokalyuk and Anton Wakolbinger, whereas Chapter 5 is joint work with Götz Kersting.

Chapter 2

Framework and synopsis of results

In the following sections we will introduce and discuss the results from Chapter 3, 4 and 5. Further, we will comment on some recent so far unpublished results obtained together with M. Birkner, I. Dahmer and C. Pokalyuk concerning the probability of fixation in populations with offspring variance tending to infinity.

Section 2.1 reviews some basic probabilistic models of the evolution of populations, including the effect of selection. In addition we examine some results in the literature concerning the survival or fixation probabilities of supercritical processes. Afterwards, in Section 2.2.1 and 2.2.2, we define neutral Cannings model admitting a paintbox construction and move on to incorporate directional selection into these models in a way that allows for a forwards and backwards description of the model. Next, in Section 2.2.3 and 2.2.4 the results and proofs of Chapter 3 and 4 are outlined. Lastly, in Section 2.2.5 we discuss the fixation probability of Cannings models with highly skewed offspring distributions, whose neutral genealogies are in the attraction of Beta-coalescents.

Section 2.3 concludes this chapter with an introduction to branching processes in an iid random environment and a discussion of the results and proofs of Chapter 5, where we prove Haldane's asymptotics for this class of processes. The proof is accomplished through an expression for the survival probability, which relies on the *shape function* for probability generating functions. We observe a connection to the theory of perpetuities, known from a financial context, and prove a limiting theorem for perpetuities with vanishing interest rates, which results in the proof of Haldane's asymptotics and a proof of a modified version of Haldane's asymptotics close the the regime of subcriticality.

2.1 Branching processes and population genetics

In this section we will summarize some essentials of the mathematical framework of theoretical population genetics that are used in this thesis. Our aim is to analyse the survival probabilities of individual-based stochastic population models, which are slightly supercritical. We call a process *slightly supercritical* if the beneficial individual has in expectation $1 + s > 1$ offspring and we are considering the limit $s \downarrow 0$. Perhaps the most basic stochastic model for an evolving population is the Galton-Watson process.

Definition 2.1.1 (Galton-Watson process).

We call $(Z_n, n \geq 0)$ a Galton-Watson process or simply a branching process, if Z_n fulfils the recursion

$$Z_n = \sum_{j=1}^{Z_{n-1}} \xi_j^{(n)}, \quad Z_0 = z_0 \geq 1,$$

where $(\xi_j^{(n)}, j \geq 1, n \geq 1)$ are independent and identically distributed on \mathbb{N}_0 , thus independent copies of a random variable ξ .

For further background on branching processes we refer to [AN72], [HJV05] and [Har63]. Galton-Watson processes are classified by the expectation $m = \mathbb{E}[\xi]$ of their offspring distribution. If the Galton-Watson process is subcritical ($m < 1$) or critical ($m = 1$), the process dies out almost surely (except in the degenerate case $\mathbb{P}(\xi = 1) = 1$). Only in the supercritical case ($m > 1$) the probability of survival is larger than 0. Letting

$$f(t) = \sum_{k=0}^{\infty} t^k \mathbb{P}(\xi = k), \quad t \in [0, 1],$$

denote the probability generating function, the survival probability $\pi = 1 - q$ can be expressed through the smallest solution q of the fixed point equation

$$q = f(q). \quad (2.1)$$

We call a sequence of Galton-Watson processes *slightly supercritical* if $m = 1 + s$ and consider the limit $s \downarrow 0$. The survival probability of a slightly supercritical Galton-Watson process with finite variance σ^2 obeys *Haldane's asymptotics*

$$\pi(s) \sim \frac{2s}{\sigma^2}, \quad \text{as } s \downarrow 0. \quad (2.2)$$

This has been shown, for example under the assumption of uniformly bounded third moments, in [Kol38], [Ath92] using the characterization of the survival probability as in (2.1). We call $s > 0$ the *selective advantage* of the beneficial individuals.

Using purely Galton-Watson processes for modelling the evolution of a population has the disadvantage that on the event of survival the population grows beyond all bounds, obviously a behaviour not observed in reality. A simple model with discrete non-overlapping generations, where the population size is kept constant, is the Wright-Fisher model [Fis30], [Wri31].

Definition 2.1.2 (Neutral Wright-Fisher model).

Fix a population size N . In the neutral Wright-Fisher model in each generation $g, g \in \mathbb{Z}$, each child chooses a parent in the prior generation uniformly and independently among the N possible parents, with the assumption that the choice of parents is independent between generations. We denote by $\nu^{(g)} = (\nu_1^{(g)}, \dots, \nu_N^{(g)})$ the number of children of the parents $1, \dots, N$ in generation g .

By the definition of the model it is clear, that ν (a generic copy of $\nu^{(g)}$) is multinomially distributed with parameters $(N; \frac{1}{N}, \dots, \frac{1}{N})$. Adding selection with parameter $s > 0$ in this model can be achieved by decreeing, that each beneficial individual now has weight $1 + s$, whereas each wildtype individual has weight 1. Each child now chooses its parent proportionally to the parent's weight. This again results in a multinomially distributed vector of offspring.

As long as beneficial individuals are rare, the expected number of offspring produced by an individual of beneficial type is approximately $1 + s$. This observation lead Haldane to argue that the survival probability $\pi(s)$ for a single beneficial mutant in the Wright-Fisher model should obey $\pi(s) \sim 2s$, for small s . Haldane approximated the binomially distributed offspring number by a Poisson distribution with parameter $1 + s$. This is in accordance with (2.2), due to $\sigma^2 = 1 = \lim_{s \downarrow 0} \text{Var}(\xi_s)$, where ξ_s is Poisson distributed with parameter $1 + s$.

While Fisher, Haldane and Wright considered models with discrete generations, more than two decades later Moran proposed a time continuous model with constant population size N and overlapping generations [Mor58].

Definition 2.1.3 (Moran model).

Assume a model with fixed population size N . For each pair of individuals $(i, j) \in [N]^2$ a reproduction event occurs at rate $\frac{1}{N}$. With probability $\frac{1}{2}$ individual i reproduces, i.e. gets an offspring and the offspring replaces individual j , with the same probability individual j reproduces.

Later on we will be concerned with Moran models whose reproduction events occur at rate $\gamma > 0$.

The Moran model can be graphically represented by considering N lines and drawing an arrow from individual i to individual j , at the moment when j gets replaced by the offspring of individual i , see Figure 2.1. In this graphical construction

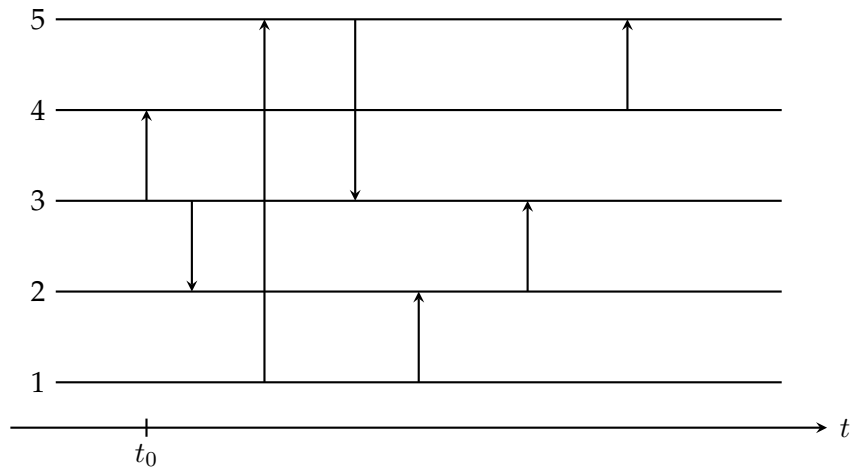


FIGURE 2.1: Graphical representation of a Moran model with 5 individuals. At time t_0 individual 4 is replaced by an offspring of individual 3.

we incorporate selection by adding selective arrows for each individual at rate s and picking the individuals to replace uniformly at random. However, only individuals of the beneficial type are able to reproduce using this mechanism, which ensures the selective advantage of the beneficial individuals, see Figure 2.2.

Another model generalising the Wright-Fisher model, now again in discrete time, is the *Cannings model* introduced by Cannings in the 1970s [Can74], [Can75], which allows for a more general offspring distribution. In order to define this class of models we recall the notion of exchangeability.

Definition 2.1.4 (Exchangeability).

We say that the random variables X_1, \dots, X_n are exchangeable, if for all permutations $\psi : [n] \rightarrow [n]$ we have

$$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\psi(1)}, \dots, X_{\psi(n)}).$$

This notion is basic in the definition of Cannings models.

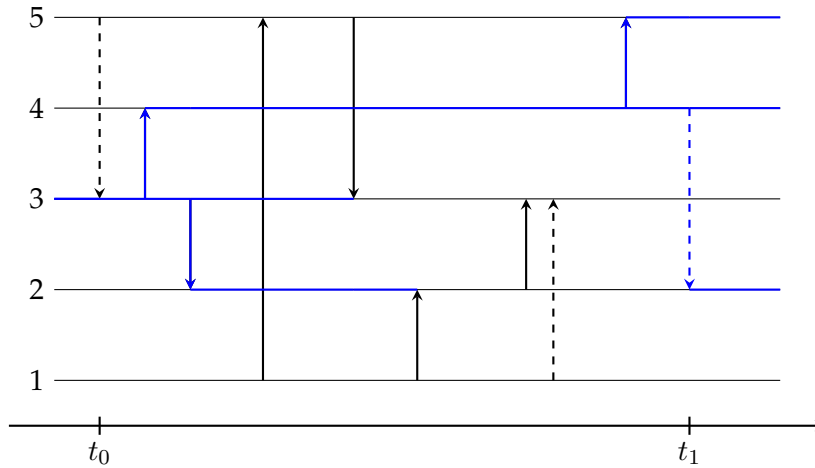


FIGURE 2.2: Selective arrows are depicted as dashed arrows. Individual 3 is of the beneficial type, shown in blue. Hence, individual 3 is not replaced at time t_0 since individual 5 is of wildtype. At time t_1 the beneficial descendant (individual 4) of individual 3 is able to use the selective arrow

Definition 2.1.5 (Cannings model).

Fix a population of size N . Let $\nu^{(g)} = (\nu_1^{(g)}, \dots, \nu_N^{(g)})$ be the vector denoting the number of offspring of the parents $1, \dots, N$ in generation g . Assume that $\mathcal{L}(\nu^{(g)})$ does not depend on the generation g and that $(\nu^{(g)}, g \in \mathbb{Z})$ are independent. Let $\nu^{(g)}$ be exchangeable, and such that $\sum_{i=1}^N \nu_i^{(g)} = N$ for all $g \in \mathbb{Z}$. Then the resulting population model is called a Canning model.

Incorporating selection into the class of Canning models is not as straightforward as in the other models, since it is not obvious how to adapt the distribution of ν in such a way that the beneficial individuals have a selective advantage. For example this has been achieved in the analytical framework of Lessard and Ladret in [LL07], however their construction did not allow for a backwards perspective on the model.

In Chapter 3 and 4 selection is incorporated to a subclass of Canning models in a way which allows to describe the *Canning frequency process* (the number of wildtype individuals in each generation) in a forwards and backwards perspective, elaborating on ideas of Krone and Neuhauser [KN97], [NK97] and of González Casanova and Spanò [GS18].

All the population models described so far in this section model the evolution of a population forwards in time, meaning that given the distribution of the types in the prior generation we are able to determine the distribution of types in the current one. In the context of mathematical population genetics it is a common approach to look backwards in time and follow the genealogy of a given sample of individuals.

Assume we are in the neutral Wright-Fisher model with two types and we aim to determine the type of a sample of n individuals. We could either look at the model forwards in time and observe the type of the n individuals through the generations or we trace the ancestors of the n individuals backwards in time and look at the type of the ancestors, which consequently gives us the type of the sample. Whenever two individuals choose the same parent their ancestral lineages coalesce, hence one speaks of *coalescent processes*.

One of the widest known coalescent processes was introduced by Kingman around the 1980s. The *Kingman coalescent* describes the time rescaled limiting genealogy e. g. of a neutral Wright-Fisher model. Let \mathcal{P}_n denote the set of all partitions of $[n]$.

Definition 2.1.6 (Kingman coalescent).

We call a \mathcal{P}_n -valued Markov process $(\Pi_n(t))_{t \geq 0}$ Kingman's coalescent, if $\Pi_n(0)$ is the partition into singletons $\{\{1\}, \dots, \{n\}\}$ and each pair of blocks merges at rate 1 independently from each other.

For neutral Cannings models the Kingman coalescent describes the limiting genealogy, whenever Möhle's condition is fulfilled [Möh00]. Möhle's condition states that if the pair coalescence probability is of smaller order than the probability of triple coalescence events

$$\frac{\mathbb{E}[\nu_1(\nu_1 - 1)]}{N - 1} = o\left(\frac{\mathbb{E}[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{(N - 1)(N - 2)}\right),$$

then the rescaled genealogy of a sample of size n converges to Kingman's coalescent. If Möhle's condition is violated other limiting genealogies are possible. Here the properly rescaled limiting genealogies are given for example as Λ -coalescents or Ξ -coalescents, which are coalescents with multiple or simultaneous collisions and were introduced independently by Pitman [Pit99], Sagitov [Sag99] and Donnelly and Kurtz [DK99]. For an overview we refer to [Ber09]. We will further comment on this case and generalisations of these models to scenarios with selection in Section 2.2.5 and study the asymptotic probability of fixation in these cases.

Related to the ideas of Kingman, which describe a neutral genealogy, are those of Krone and Neuhauser in 1997, who introduced the *ancestral selection graph* for the Moran model with selection [KN97] and [NK97]. The ancestral selection graph in the Moran model with selection describes the genealogy of a sample backwards in time and briefly stated the idea goes as follows.

Looking backwards in time the types of the parent individuals are not known. Therefore, if we trace a lineage backwards in time, it is not clear which selective arrows have actually been used for reproduction. Hence, whenever there is a selective arrow both *potential parents* are traced further backwards in time which results in a branching event for the ancestral process. Whenever there is a neutral reproduction event among the individuals in the ancestral process, we observe a coalescing event as in Kingman's coalescent. Note that an individual is of beneficial type if and only if at least one of its potential parents is of beneficial type.

In the following we will be concentrating on the *ancestral selection process*, which counts the number of blocks in the ancestral selection graph.

Definition 2.1.7 (Moran ancestral selection process).

Assume the setting of a Moran model with selection strength $s > 0$. We call the Markov process $\mathcal{B} = (B_r)_{r \geq 0}$ with state space $[N]$ the ancestral selection process, if its transition rates $r_{x,y}$ for $x \in [N]$ are given as

$$r_{x,x+1} = xs \left(1 - \frac{x}{N}\right),$$

$$r_{x,x-1} = \binom{x}{2} \frac{\gamma}{N},$$

for some parameter $\gamma > 0$, and $r_{x,y} = 0$ for all $y \notin \{x - 1, x + 1\}$.

Note that the Moran ancestral selection process with parameter γ corresponds to a Moran model whose neutral reproduction events occur at rate γ for each pair of individuals.

The results of Chapter 3 and 4 are twofold in the following manner. On the one hand we introduce a graphical construction for Cannings models, in the spirit of the graphical construction of the Moran model, which allows to incorporate selection with strength $0 < s < 1$ into a subclass of Cannings's models. This "forward" construction is reminiscent of the way selection is added into the Wright-Fisher model, but is flexible enough to allow for many possible limiting genealogies, including Λ -coalescents or Ξ -coalescents.

On the other hand the way selection is introduced allows a backwards description of the Cannings models with selection, which enables us to define a *Cannings ancestral selection graph* and the corresponding *Cannings ancestral selection process*. The model and both processes are defined in Section 2.2.1 and Section 2.2.2.

With the help of the Cannings ancestral processes we are able to prove Haldane's asymptotics for the case of moderately weak selection in Cannings models whose neutral genealogies are in the attraction of Kingman's coalescent. We speak of the regime of *moderate selection* with selective strength $0 < s_N < 1$, if

$$s_N \rightarrow 0, \quad s_N N \rightarrow \infty, \quad \text{as } N \rightarrow \infty.$$

See e.g. [PW20] where the case of moderate selection was considered in a related setting.

Another regime of selection studied in population genetics is the regime of *weak selection*, where

$$s_N \sim \frac{\alpha}{N}, \quad \text{as } N \rightarrow \infty,$$

for some $\alpha > 0$. Here selection and the coalescence probability c_N are of the same order, i.e. $s_N = \Theta(c_N)$, which is a different regime than the ones considered in Section 2.2.3 to 2.2.5. In the case of weak selection for large N the frequency of beneficial individuals can be approximated by a *Wright-Fisher diffusion* with selection,

$$dX_t = \alpha dt + X_t(1 - X_t)dW_t, \quad X_0 = p \in (0, 1)$$

where $(W_t)_{t \geq 0}$ is an independent Wiener process, see e.g. [Eth11]. In the diffusion approximation the probability of fixation π can be calculated explicitly and is given by

$$\pi = \frac{1 - e^{-\frac{2\alpha p}{\sigma^2}}}{1 - e^{-\frac{2\alpha}{\sigma^2}}}, \quad (2.3)$$

where $p \in (0, 1)$ denotes the initial frequency of the beneficial individuals and in the classical Wright-Fisher model we have $\sigma^2 = 1$. Again note that (2.3) is an increasing function in α and a decreasing function in σ , which shows the same interplay of genetic drift and selection as in Haldane's asymptotics. The case of weak selection in large populations was among others considered by Gillespie [Gil74], [Gil75], Kimura [Kim62] and Moran [Mor59], [Mor60]. They concluded that the fixation probability π_N in a large population of size N in the case of weak selection is asymptotically given by (2.3). This asymptotics has also been shown with the help of ancestral processes by Pfaffelhuber and Pokalyuk in [PP13], who studied

the equilibrium distribution of the Moran ancestral selection process with weak selection. To the present day ancestral processes have been used to study a wide range of problems related to population genetics, not only concerning models with selection. This includes models which incorporate mutations [BLW16], [BCH18], [LKBW15] or [Tay07], recombination [BB20], [GM97], or with diploid populations [BBE13], [BLS18], also see [BW21].

At this point let us further mention some results connected to the topic of selection acting on a population. The long term evolutionary experiment performed by Lenski et al. since 1988 (see [LRST91] and [LT94]) studies an evolving population of *Escherichia coli* under selective pressure. Every day the bacteria were given a nutrient solution such that they could reproduce and at the end of the day approximately one percent of the bacteria was sampled as a founding colony for the next day. This experiment was the motivation for the papers [GKWY17] and [BGPW19]. The sampling procedure of bacteria by Lenski et al. is reminiscent of the sampling procedure in Cannings models as in [Sch03] and was modelled in [GKWY17] in the following way. In every cycle, each individual produces a random number of potential offspring. From this pool a random sample of size N is taken without replacement, which then found the next generation. Hence, as we will see in Section 2.2.1, Cannings models admitting a paintbox construction can be seen as the sampling with replacement analogue (and generalisation of the offspring distribution) of the population model developed in [GKWY17].

In this thesis we do not study the situation where multiple beneficial or deleterious mutations can accumulate as in [EPW09] and [DM11] (also see [RBW08] and [DF07]). In [EPW09] one considers the question how fast deleterious mutations are accumulated over time and analyses the time it takes the fittest type to vanish. The situation in [DM11] is comparable, they consider the questions how fast beneficial mutations accumulate and hence they analyse the speed at which the population evolves, measured by the fitness difference of the current population compared to the starting fitness. It might be promising for further research to link this framework with the model described in the following section.

2.2 Haldane's asymptotics for Cannings models with moderate selection

2.2.1 A graphical representation including selection

Consider a classical Cannings model [Can74], [Can75] with fixed population size $N \in \mathbb{N}$, which is characterized by the distribution of its exchangeable and nonnegative offspring vector $\nu = (\nu_1, \dots, \nu_N)$, such that $\nu_1 + \dots + \nu_N = N$, see Definition 2.1.5. In the following we will concentrate on the special subclass of Cannings models which admit a *paintbox construction*.

We introduce the vector $\mathscr{W}^{(N)} = (W_1^{(N)}, \dots, W_N^{(N)})$ of random weights, such that $\mathscr{W}^{(N)}$ is exchangeable and

$$\sum_{j=1}^N W_j^{(N)} = 1, \quad W_i^{(N)} \geq 0, \quad i \in [N].$$

For notational simplicity we do not always display dependence of $\mathscr{W}^{(N)}$ on N , and simply write \mathscr{W} . However, sometimes we want to emphasize the random weights belonging to the individuals in generation g , which we denote by $\mathscr{W}^{(g)}$. We assume that the family $(\mathscr{W}^{(g)})_{g \in \mathbb{Z}}$ is independent and identically distributed.

The number of offspring from one generation to the next is generated by a two-step procedure. First, sample the vector of random weights \mathscr{W} , then given \mathscr{W} we let ν be multinomially distributed with parameters $(N; \mathscr{W})$. For example, one obtains the classical Wright-Fisher model (Definition 2.1.2), if one chooses $W_i = \frac{1}{N}$ for all $i \in [N]$. A special class of random weights, the class of *Dirichlet-type* weights, will be introduced in Section 2.2.4.

Let us briefly recall the parental relations between individuals in generation $g-1$ and g . Denote by (i, g) , $i \in [N]$ the individuals in generation g . Each individual (i, g) is assigned a parent $(V(i, g), g-1)$ in generation $g-1$, with the property that for all $j \in [N]$

$$\mathbb{P}\left(V(i, g) = j \mid \mathscr{W}^{(g-1)}\right) = W_j^{(g-1)}, \quad (2.4)$$

conditionally on $\mathscr{W}^{(g-1)}$ independently for all $i \in [N]$ and independently over the generations. An advantage of this construction is the existence of the following graphical representation to construct the parental relations in this class of Cannings models.

For every $g \in \mathbb{Z}$ let $U_1^{(g)}, \dots, U_N^{(g)}$ be independent and uniformly distributed random variables on the unit square $[0, 1] \times [0, 1]$ and divide the square into N stripes $[0, 1] \times \mathscr{I}_1^{(g-1)}, \dots, [0, 1] \times \mathscr{I}_N^{(g-1)}$, such that $\mathscr{I}_j^{(g-1)}$ has length $W_j^{(g-1)}$, precisely

$$\mathscr{I}_j^{(g-1)} = \left[\sum_{i=1}^{j-1} W_i^{(g-1)}, \sum_{i=1}^j W_i^{(g-1)} \right], \quad j \in [N].$$

In light of (2.4), we have the equality of events

$$\{V(i, g) = j\} = \{U_i^{(g)} \in [0, 1] \times \mathscr{I}_j^{(g-1)}\},$$

which is illustrated in Figure 2.3.

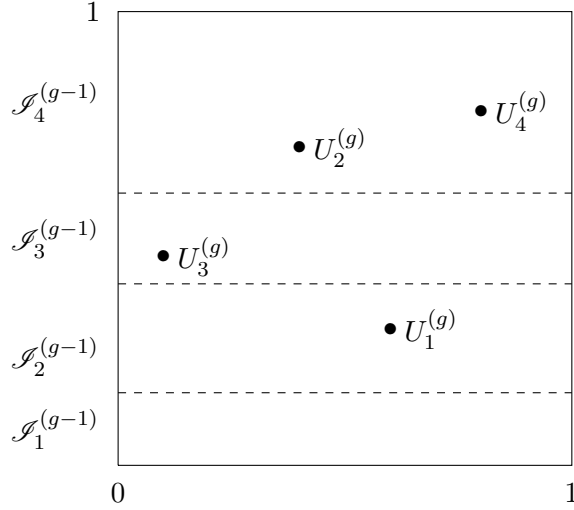


FIGURE 2.3: A realisation of a neutral Cannings model admitting a paintbox representation with $N = 4$ individuals. For example, individual 2 in generation g has chosen individual 4 in generation $g - 1$ as its parent, since $U_2^{(g)} \in [0, 1] \times \mathcal{I}_4^{(g-1)}$.

We incorporate directional selection with strength $s_N \in (0, 1)$ into this model in the following way. Assume each individual is either of *wildtype* or *beneficial type* and each child inherits the type of its parent. Define the vector of modified weights $\widetilde{\mathcal{W}}^{(g-1)} = (\widetilde{W}_1^{(g-1)}, \dots, \widetilde{W}_N^{(g-1)})$ such that $\widetilde{W}_i^{(g-1)} = (1 - s_N)W_i^{(g-1)}$, if i is of wildtype and $\widetilde{W}_i^{(g-1)} = W_i^{(g-1)}$, if i is of beneficial type, for all $i \in [N]$. Hence, the weight of each wildtype individual is reduced by the factor $(1 - s_N)$, resulting in an approximate selective advantage of $1 + s_N$ for a single beneficial individual.

In the same manner as before, generation g results from a multinomial sampling with weights proportional to $\widetilde{\mathcal{W}}^{(g-1)}$. More precisely, we decree that $(j, g - 1)$ is the parent of individual (i, g) with probability

$$\mathbb{P}\left(V(i, g) = j \mid \widetilde{\mathcal{W}}^{(g-1)}\right) = \frac{\widetilde{\mathcal{W}}_j^{(g-1)}}{\sum_{k=1}^N \widetilde{\mathcal{W}}_k^{(g-1)}}, \quad j \in [N],$$

again conditionally on $\widetilde{\mathcal{W}}^{(g-1)}$ independently for all $i \in [N]$ and independently between the generations.

The sampling procedure can be incorporated into the graphical construction by defining the sets

$$\begin{aligned} \mathcal{B}^{(g-1)} &:= \{i \in [N] : (i, g - 1) \text{ is of beneficial type}\}, \\ \mathcal{C}^{(g-1)} &:= \{i \in [N] : (i, g - 1) \text{ is of wildtype}\}. \end{aligned}$$

Now let $(U_1^{(g)}, \dots, U_N^{(g)})$ be independent and uniformly distributed on $\Gamma^{(g-1)}$, with

$$\Gamma^{(g-1)} := \bigcup_{j \in \mathcal{B}^{(g-1)}} [0, 1] \times \mathcal{I}_j^{(g-1)} \cup \bigcup_{j \in \mathcal{C}^{(g-1)}} [0, 1 - s_N] \times \mathcal{I}_j^{(g-1)}.$$

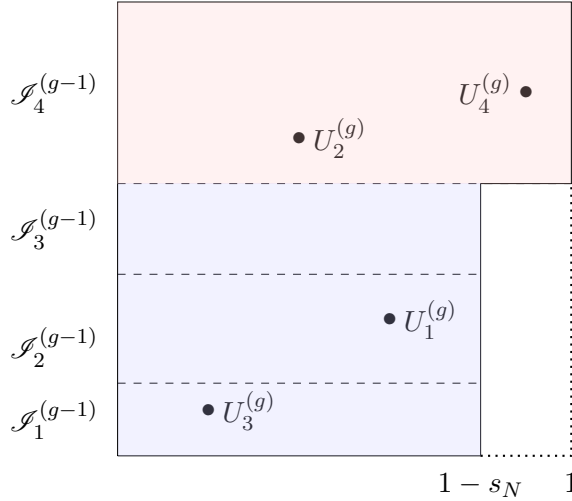


FIGURE 2.4: An example of a realisation of a transition step in a Cannings model with $N = 4$ individuals. Individual 4 is beneficial, whereas the other three individuals are of wildtype.

We then have the equality of events (compare with (2.4) and see Figure 2.4),

$$\begin{aligned} \{V(i, g) = j\} &= \{U_i^{(g-1)} \in [0, 1] \times \mathcal{J}_j^{(g-1)}\}, & \text{for } j \in \mathcal{B}^{(g-1)}, \\ \{V(i, g) = j\} &= \{U_i^{(g-1)} \in [0, 1 - s_N] \times \mathcal{J}_j^{(g-1)}\}, & \text{for } j \in \mathcal{C}^{(g-1)}. \end{aligned}$$

In the following we aim to study the extinction of the wildtype individuals (or equivalently the fixation of the beneficial individuals). To keep track of their frequency along the generations we define their counting process.

Definition 2.2.1 (Cannings frequency process).

Denote by K_g the number of wildtype individuals in generation g and denote by $\mathcal{K} = (K_g)_{g \geq 0}$ the Cannings frequency process with selection strength s_N , population size N and weight distribution $\mathcal{L}(\mathcal{W})$.

Note that due to the exchangeability of \mathcal{W} the transition probabilities of \mathcal{K} do not depend on the precise configuration of beneficial individuals in the prior generation, hence \mathcal{K} is a Markovian process. Furthermore, due to the sampling procedure as described before, given $K_{g-1} = k$ and $\mathcal{W}^{(g-1)}$ the transition probabilities are the same as a mixed binomial random variable with parameters N and $P(k, \mathcal{W})$, with

$$P(k, \mathcal{W}) := \frac{(1 - s_N) \sum_{i=1}^k W_i}{(1 - s_N) \sum_{i=1}^k W_i + \sum_{i=k+1}^N W_i}.$$

The Cannings frequency process describes the evolution of the wildtype individuals forwards in time, however the aforementioned construction also allows for a backwards description of the model, which is introduced in the next section.

2.2.2 The Cannings ancestral selection process

Due to the construction of Cannings models as in Section 2.2.1, where the offspring distribution results from a paintbox sampling procedure, there exists an ancestral process \mathcal{A} , which is dual to \mathcal{K} . The *Cannings ancestral selection process* $\mathcal{A} = (A_m)_{m \in \mathbb{N}_0}$ with parameters N, s_N and $\mathcal{L}(\mathcal{W})$ counts the number of potential parents from any subset of individuals m generations in the past.

The Cannings ancestral selection process is a Markov chain with state space $[N]$, whose transition probabilities are governed by a branching and coalescing step. With $A_m = a$, first in the branching step generate a random sum $H = \sum_{i=1}^a G^{(i)}$, with $G^{(i)}$ being independent geometric random variables with parameter $1 - s_N$, thus H is negatively binomial distributed. Next, for the coalescing step, take a vector of random weights $\mathcal{W} = (W_1, \dots, W_N)$ and put H balls randomly into N boxes, where box i has weight W_i . The distribution of A_{m+1} is the same as the distribution of the number of occupied boxes after this two step procedure.

Let us elaborate how this procedure arises naturally and explain why the distribution of A_{m+1} given $A_m = a$ is described by this two-step procedure. A priori there is no information given on how many beneficial individuals are present in the prior generation. Therefore, we need to ensure that for every possible configuration of beneficial individuals in the prior generation, the parental relations between the generations are distributed accordingly. Hence, we split each weight W_i into a neutral part of mass $(1 - s_N)W_i$ and a selective part of mass $s_N W_i$. This leads to the square $[0, 1] \times [0, 1]$ being vertically divided into a neutral part \mathcal{N} and a selective part \mathcal{S} ,

$$\mathcal{N} := [0, 1 - s_N] \times [0, 1] \quad \mathcal{S} := [1 - s_N, 1] \times [0, 1],$$

as depicted in Figure 2.5. For each individual $j \in [N]$ in generation g consider a sequence $U_j^{(1,g)}, U_j^{(2,g)}, \dots$ of independent and uniform distributed random variables on the unit square. We now proceed as follows, for each individual we wait until the first of these uniforms falls into the area \mathcal{N} and denote this number by $\gamma(j, g)$,

$$\gamma(j, g) := \min\{\ell \geq 1 : U_j^{(\ell,g)} \in \mathcal{N}\}.$$

This ensures no matter the configuration of beneficial individuals in the prior generation, that there will be a valid parent of (j, g) . Obviously, $\gamma(j, g)$ is geometrically distributed with parameter $(1 - s_N)$ and we call the individuals $k \in [N]$ such that

$$U_j^{(\ell,g)} \in [0, 1] \times \mathcal{I}_k^{(g-1)}, \quad \text{for some } 1 \leq \ell \leq \gamma(j, g)$$

potential parents of (j, g) . See Figure 2.5 for an illustration.

Since we started with a individuals we obtain through this procedure a different sets of potential parents. We are only keeping track of the total number of distinct potential parents among these a sets, which corresponds to checking how many stripes $\mathcal{I}_k^{(g-1)}$ are occupied. This leads to the aforementioned coalescing step where H balls are put randomly into N boxes with weights (W_1, \dots, W_N) .

Observe that an individual is of wildtype if and only if all its potential parents are of wildtype. Put differently, an individual is of beneficial type if and only if at least one of its potential parents are of beneficial type. This observation is key for the following sampling duality, connecting the forward and backward processes $\mathcal{K} = (K_g)_{g \in \mathbb{N}_0}$ and $\mathcal{A} = (A_g)_{g \in \mathbb{N}_0}$.

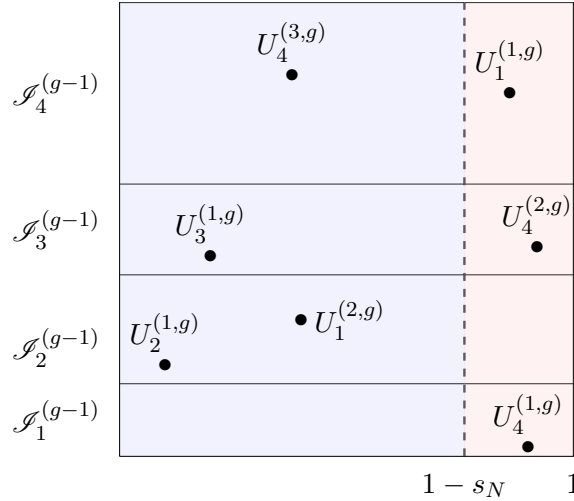


FIGURE 2.5: A realisation of potential parents in a Cannings model with $N = 4$ individuals. Individual 1 and 4 have chosen two respectively three potential parents, with $\gamma(1, g) = 2$ and $\gamma(4, g) = 3$. Individuals 2 and 3 each have a single potential parent. E.g. individual 3 and 4 have in total three potential parents.

Theorem 1 (Sampling Duality).

Let $g \geq 0$ and $k, n \in [N]$, then the following duality relations holds

$$\mathbb{E} \left[\frac{K_g(K_{g-1}) \cdots (K_g - n + 1)}{N(N-1) \cdots (N-n+1)} \mid K_0 = k \right] = \mathbb{E} \left[\frac{k(k-1) \cdots (k - A_g + 1)}{N(N-1) \cdots (N - A_g + 1)} \mid A_0 = n \right].$$

The above duality can be interpreted in the following way. Given there are k wildtype individuals at time 0, the probability that a sample of size n at time g only consists of wildtype individuals is the same as the probability that all A_g potential ancestors of the sample are among the k wildtype individuals. Note that the same duality holds true in the Moran model with selection for the Moran frequency process and the Moran ancestral selection process, but in continuous time instead, see Section 3.4 and compare [AS05], [Man09].

Specializing Theorem 1 to the case $k = N - 1$ and $n = N$ yields an useful expression for the fixation probability of a single beneficial mutant, which is used in the proof of Theorem 2.

Corollary 2.2.2. Let A_{eq} have the stationary distribution of the Cannings ancestral selection process. Then we have

$$\pi_N := \lim_{g \rightarrow \infty} \mathbb{P}(K_g = 0 \mid K_0 = N - 1) = \mathbb{E} \left[\frac{A_{eq}}{N} \right].$$

Note that Theorem 1 allows to express the survival probability for any number of initially present beneficial individuals. However, for this it is necessary to control higher moments of A_{eq} . So far the distribution of A_{eq} is not explicitly known, therefore it seems rather burdensome to accomplish sufficient control of the higher moments of A_{eq} . In the case of the Moran ancestral selection process the equilibrium distribution is known and there it is possible to calculate the survival probability, if one starts with $k \in [N]$ beneficial individuals, see Section 3.4.

2.2.3 The case of moderately weak selection

We say that a Cannings model with directional selection with strength $s_N \in (0, 1)$ is in the regime of *moderately weak selection*, if

$$N^{-1+\eta} \leq s_N \leq N^{-\frac{1}{2}-\eta}, \quad (2.5)$$

for some $\eta > 0$. Often we equivalently denote $b = b_N := -\frac{\log s_N}{\log N}$, where (2.5) translates to $\frac{1}{2} < b < 1$.

Throughout this section we assume, that for some $\rho^2 \geq 1$ we have

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^2 \right] = \frac{\rho^2}{N^2} + O(N^{-3}), \quad (2.6)$$

as $N \rightarrow \infty$. Note that the requirement $\rho^2 \geq 1$ is no restriction, due to $\text{Var} \left(W_1^{(N)} \right) \geq 0$. Note that assumption (2.6) yields for the variance of the offspring distribution

$$\text{Var} \left(\nu_1^{(N)} \right) = \rho^2 + o(1), \quad \text{as } N \rightarrow \infty.$$

The main result of Chapter 3 is proven under two different sets of assumptions.

Assumptions A1: We consider the following condition on the third moment of $W_1^{(N)}$ and the smaller regime of moderately weak selection

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^3 \right] = O(N^{-3}), \quad N^{-1+\eta} \leq s_N \leq N^{-\frac{2}{3}-\eta}, \quad (2.7)$$

for some $\eta > 0$, which corresponds to $\frac{2}{3} < b < 1$.

Alternatively, under **Assumptions A2** we assume that there exists a sequence of natural numbers h_N , such that

$$h_N \rightarrow \infty \quad \text{and} \quad h_N = o(\log N), \quad \text{as } N \rightarrow \infty,$$

with the property that for N big enough we have

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^n \right] \leq \left(\frac{K h_N}{N} \right)^n, \quad (2.8)$$

for some constant $K \geq 1$ and $n \leq 2h_N$. Furthermore, assume that s_N is as in (2.5).

Remark 2.2.3. Note that (2.7) as well as (2.8) together with (2.6) imply

$$N \mathbb{E} \left[\left(W_1^{(N)} \right)^2 \right] \rightarrow 0, \quad \text{and} \quad \mathbb{E} \left[\left(W_1^{(N)} \right)^3 \right] = o \left(\mathbb{E} \left[\left(W_1^{(N)} \right)^2 \right] \right),$$

which is sufficient for the neutral genealogy of a Cannings model with a paintbox construction to be in attraction of Kingman's coalescent [Kin82b], as a consequence of Möhle's Lemma, see [Möh00].

Theorem 2 (Haldane's formula, moderately weak selection).

Let $(K_g^{(N)})_{g \geq 1}$ be a sequence of Cannings frequency processes, with parameters N, s_N and $\mathcal{L}(\mathcal{W}^{(N)})$. Assume that (2.6) is fulfilled. Then under either assumptions A1 or A2 we have for the fixation probability π_N of a single beneficial mutant, that

$$\pi_N \sim \frac{2s_N}{\rho^2} \quad \text{as } N \rightarrow \infty.$$

The proof of Theorem 2 takes advantage of the duality in Corollary 2.2.2, which allows to express the fixation probability in terms of the equilibrium distribution of the Cannings ancestral selection process. However, approximating the expected value of the Cannings ancestral selection process in stationarity turns out to be challenging. In Chapter 3 we lay out two different strategies of proof, under Assumptions A1 or A2 respectively. Assumption A2 requires control of higher moments of $W_1^{(N)}$ compared to A1, however we are able to cover the whole regime of moderately weak selection.

Under assumptions A1 a coupling approach with the Moran ancestral selection process $\mathcal{B} = (B_t)_{t \geq 0}$ with parameter $\gamma = \rho^2$ is successful, for a precise definition see Definition 2.1.7. The stationary distribution of \mathcal{B} is a binomial distribution with parameters $(N, \frac{2s_N}{\rho^2 + 2s_N})$ conditioned to be strictly positive, [Cor17] and [CM19]. One easily checks that

$$\mathbb{E}[B_{eq}] \sim \frac{2s_N N}{\rho^2},$$

with B_{eq} being binomially distributed with parameters $(N, \frac{2s_N}{\rho^2 + 2s_N})$ conditioned to be strictly positive. Therefore, we need to ensure that the processes in stationarity $(A_g^{eq})_{g \in \mathbb{N}_0}$ and $(B_t^{eq})_{t \geq 0}$ can be coupled sufficiently well, such that

$$\mathbb{E}[A_{eq}] = \mathbb{E}[B_{eq}] (1 + o(1)) \sim \frac{2s_N N}{\rho^2}, \quad (2.9)$$

where A_{eq} is distributed according to the stationary distribution of $(A_g)_{g \geq 0}$. Due to Corollary 2.2.2 this yields Haldane's asymptotics.

We achieve this coupling by considering three processes \mathcal{A}, \mathcal{B} and \mathcal{B}_{eq} , where \mathcal{B}_{eq} is a Moran ancestral selection process started in equilibrium. In the spirit of Döblin, we couple \mathcal{B} and \mathcal{B}_{eq} such that these two processes coincide once they simultaneously hit the same state.

For the purpose of coupling \mathcal{A} and \mathcal{B} such that (2.9) holds, we identify a neighbourhood U of $\frac{2s_N N}{\rho^2}$, such that \mathcal{B} enters this neighbourhood sufficiently fast and stays in U until it has approached its stationary distribution close enough. Within U the processes \mathcal{A} and \mathcal{B} can be controlled to remain close for a sufficiently long time. Precisely, due to Lemma 3.6.8 we get that for any starting distribution of \mathcal{B} after a time of order $N^{b+\delta}$, $\delta > 0$ the process is close enough to its stationary distribution such that

$$\mathbb{E}[B_{N^{b+\delta}}] \sim \mathbb{E}[B_{eq}] \sim \frac{2s_N N}{\rho^2}.$$

Therefore, we chose U such that \mathcal{A} and \mathcal{B} stay in U for a time of order $N^{b+\delta}$ with high probability. This is achieved in Lemma 3.6.7 for \mathcal{B} . In Lemma 3.6.1 we observe that within U both processes have similar transition probabilities, thus in Lemma 3.6.9 we show that \mathcal{A} and \mathcal{B} can be coupled for a time of order $N^{b+\delta}$ such that with high probability

$$|A_g - B_g| \leq 1, \quad \text{for } 0 \leq g \leq N^{b+\delta},$$

if \mathcal{A} is started in U . Furthermore, we ensure that both processes enter the region U quickly enough, which is shown in Lemma 3.6.6 and Lemma 3.6.3. Altogether, this allows us to deduce the desired asymptotics for the expectations in (2.9).

Under Assumptions A2 a different strategy of proof turns out to be successful.

We estimate the drift of A_{eq} in every state $x \in [N]$ and observes that there is a unique center of attraction at $\frac{2s_N N}{\rho^2}(1 + o(1))$, meaning that in expectation the process moves towards $\frac{2s_N N}{\rho^2}(1 + o(1))$. Similar as before we define a neighbourhood U' around $\frac{2s_N N}{\rho^2}$ which is large enough such that A_{eq} remains in U' long enough, but on the other hand not too big such that we can deduce

$$\mathbb{E}[A_{eq}] \sim \frac{2s_N N}{\rho^2}, \quad (2.10)$$

which by Corollary 2.2.2 yields Haldane's asymptotics. More precisely, in Lemma 3.7.4 and Lemma 3.7.5 we show that the process returns fast enough to U' , that is in a polynomially long time with exponential high probability. Complementary, Proposition 3.7.3 shows that the process stays in U' for an exponentially long time, with sufficiently high probability. To prove the latter, a comparison argument to a random walk with drift towards $\frac{2s_N N}{\rho^2}$, is applied. Altogether, this proves (2.10).

Remark 2.2.4. *Let us briefly discuss the two arising boundary cases $b = \frac{2}{3}$ and $b = \frac{1}{2}$ under the assumptions A1 and A2.*

The strategy of proof under assumptions A1 in Section 3.6 relies on the coupling of \mathcal{A} and \mathcal{B} in the region U . This is a rather sharp coupling, in the sense that only a difference of 1 between A_g and B_g is allowed for a time interval, which length is of order $N^{b+\delta}$, $\delta > 0$. However, such a coupling can only hold if jumps of size 2 or larger are unlikely in U , which is the statement of Lemma 3.6.1. Below the boundary $b = \frac{2}{3}$ the behaviour of \mathcal{A} in U changes such that jumps of arbitrary size become possible, which can be seen by the following heuristic.

Let $a_{eq} = a_{eq}^{(N)} = \frac{2s_N N}{\rho^2}$ denote our candidate for the asymptotic expectation under stationarity of \mathcal{A} . The probability that two pairs coalesce in a single transition step of \mathcal{A} is asymptotically given as $\frac{a_{eq}^4}{N^2}$, the squared pair-coalescence probability. Since the coupling needs to hold for $N^{b+\delta}$ generations we require

$$N^{b+\delta} \frac{a_{eq}^4}{N^2} \sim cN^{2-3b+\delta} \rightarrow 0,$$

which entails $b > \frac{2}{3}$. Thus, the jumps of \mathcal{A} at the central region become larger and our coupling approach fails.

To circumvent this problem and to prove Haldane's asymptotics for $b > \frac{1}{2}$ we apply a different technique of proof under Assumptions A2. We observe that near the assumed expectation of A_{eq} the process behaves comparable to a random walk with drift. For this argument to work successfully it is essential that the jump sizes of this random walk in the region U' are bounded with high probability. This is shown via Lemma 3.7.1 for the upward jumps and in Lemma 3.7.2 for the downward jumps, which both rely on the assumption $b > \frac{1}{2}$. Let a_{eq} be as above. The branching part of the transition step of \mathcal{A} transforms a_{eq} into the random sum $H = \sum_{i=1}^{a_{eq}} G_i$ with G_i being independent geometric random variables with parameter $1 - s_N$. Hence, the expected number of additional branches is given as

$$\mathbb{E}[H] - a_{eq} = a_{eq} \left(\frac{1}{1 - s_N} - 1 \right) \sim s_N a_{eq}.$$

The above entails that in expectation we get N^{1-2b} additional branches, letting $b < \frac{1}{2}$ the number of additional branches diverges as $N \rightarrow \infty$, which breaks our coupling with a random walk with bounded jump sizes. The same difficulty arises for the downward jumps.

Hence, at both boundaries the transition probabilities of \mathcal{A} change such that larger jumps become likely, at the boundary $b = \frac{2}{3}$ jumps of size larger than 2 become likely, but jump sizes

stay bounded with high probability and at the boundary $b = \frac{1}{2}$ even larger jumps become possible. Consequently, a coupling approach becomes harder to realise at these boundaries.

2.2.4 The case of moderately strong selection

Similarly as in Section 2.2.3 we are again considering the regime of moderate selection. However, selection is now on the stronger side, i.e. there exists $\eta > 0$ such that

$$N^{-\frac{1}{2}+\eta} \leq s_N \leq N^{-\eta}, \quad (2.11)$$

which connects well to the regime of moderately weak selection in (2.5). With the same notion as before, $b = b_N = -\frac{\log s_N}{\log N}$ the above reads $0 < b < \frac{1}{2}$.

In this section we assume that the weights $\mathscr{W}^{(N)}$ are of *Dirichlet-type*.

Definition 2.2.5 (Dirichlet-type weights).

We say that a vector of random exchangeable weights $\mathscr{W}^{(N)}$ is of *Dirichlet-type* if there exist Y_1, \dots, Y_N which are independent copies of Y with $\mathbb{P}(Y > 0) = 1$, such that

$$W_i^{(N)} = \frac{Y_i}{\sum_{j=1}^N Y_j}, \quad i \in [N].$$

For example, if Y is gamma distributed, \mathscr{W} is symmetrically Dirichlet distributed. Furthermore, the Wright-Fisher model corresponds to $Y = c$ almost surely, for some $c > 0$.

The main theorem in Chapter 4 is proven under the assumption

$$\mathbb{E}[\exp(hY)] < \infty, \quad (2.12)$$

for some $h > 0$. For a possible relaxation of this condition see Remark 4.3.2 a) and 4.3.3 a). For example in the case of slowly varying tails, one can weaken this assumption to the existence of the 4-th moment. Assumption (2.12) implies by Möhle's Lemma that the neutral genealogy is in attraction of Kingman's coalescent, see Lemma 4.4.2.

The main result of Chapter 4 is that for Cannings models with moderately strong selection and Dirichlet-type weights fulfilling (2.12), Haldane's asymptotics holds true as well.

Theorem 3 (Haldane's formula, moderately strong selection).

Let $(K_g^{(N)})_{g \geq 1}$ be a sequence of Cannings frequency processes with parameters N, s_N and $\mathscr{L}(\mathscr{W}^{(N)})$ and assume that the random weights are of Dirichlet-type fulfilling (2.12) and that (2.11) and (2.6) hold. Then the fixation probability π_N of a single beneficial mutant obeys

$$\pi_N \sim \frac{2s_N}{\rho^2} \quad \text{as } N \rightarrow \infty.$$

Let us briefly discuss the strategy of proof. The proof of Theorem 3 is divided into three phases, in which the number of beneficial individuals reaches the levels $N^{b+\delta}, \varepsilon N, N$ respectively, for some $\varepsilon, \delta > 0$.

In the first phase, to reach the level $N^{b+\delta}$ the number of beneficial individuals is approximated from above and below by a Galton-Watson process with a mixed Poisson offspring distribution and a mixed Binomial offspring distribution, respectively, see Lemma 4.5.1. Both offspring distributions have offspring mean $1 + s_N + o(s_N)$ and offspring variance $\rho^2 + o(1)$. Relying on results by ([Ath92], [Esh81] or [Hop92]) on

the survival probability of slightly supercritical Galton-Watson processes, we show that the number of beneficial individuals reaches the level $N^{b+\delta}$ with probability $\frac{2s_N}{\rho^2}(1 + o(1))$.

For the second phase, the frequency process of the beneficial individuals grows in a first order approximation deterministically, where in each generation the process grows by at least the factor $(1-\varepsilon)s_N$ until it reaches the level εN with high probability. Let g_0 be the number of generations needed, such that the expected number of beneficial individuals reaches the level $2\varepsilon N$. We estimate the accumulated variance after g_0 generations for this process, see Lemma 4.5.4 and observe that the variance is of smaller order than the squared expected value after g_0 steps. Hence, we deduce that the level εN is reached with high probability.

Lastly, we make use of the duality (Theorem 1) and the fact that the wildtype fixates if and only if there is no beneficial individual in the set of potential ancestors. However, the number of potential ancestors in equilibrium converges to ∞ in probability, as $N \rightarrow \infty$. Hence, if initially at least εN (out of N) individuals are of beneficial type, the probability that all potential ancestors are of wildtype converges to 0, therefore the beneficial type fixates with high probability.

Remark 2.2.6. *a) As in the case of moderately weak selection we discover the boundary case $b = \frac{1}{2}$. The most crucial step of the proof of Theorem 3 is the approximation with slightly supercritical Galton-Watson processes. If the number of beneficial individuals reaches the level $N^{b+\delta}$, they fixate with probability tending to 1. Hence, to go beyond $b = \frac{1}{2}$ the approximation with Galton-Watson processes has to hold for more than \sqrt{N} individuals. However, at the level \sqrt{N} , due to the birthday problem, the beneficial individuals start to interfere. With non-trivial probability in the lower bounding Galton-Watson process a new beneficial individual is added, while in the original process a beneficial individual replaces another beneficial individual. Consequently, the approximation with a Galton-Watson process from below fails.*

This phenomenon has been observed also by other authors modelling interacting systems with the help of branching processes ([BR13], [BD95], [GKWY17]). Nevertheless, we have no reason to doubt that Haldane's asymptotics holds true in this case. It seems plausible that a connection of the forwards and backwards approaches from Chapter 3 and Chapter 4 might be successful. For another possible route to attack $b = \frac{1}{2}$ see Remark 2.2.9.

b) The application of the duality (Theorem 1) allows to replace and simplify the commonly used approach to show fixation of the beneficial individuals after the number of beneficial individuals has reached the level εN . Usually, this approach involves two steps, first an approximation with ordinary differential equation is used, in the regime where the frequency process reaches the level $(1-\varepsilon)N$ from εN and afterwards a comparison of the wildtype individuals to a subcritical Galton-Watson process, which consequently dies out with probability 1, see for example [GKWY17].

At this point we can give some motivation for the questions in [BK21], which are connected to the proof of Theorem 3. As discussed an extension of the proof of Theorem 3 to a regime $b \geq \frac{1}{2}$ is not straightforward. The proof relies on the coupling with Galton-Watson processes until the Galton-Watson process reaches a level of order $N^{b+\delta}$, $\delta > 0$. This coupling crucially depends on Lemma 4.4.4, which allows us to consider the random weights as if they were independent for $k \leq N^{b+\delta}$ up to some small correction and therefore we can compare the Cannings frequency process to Galton-Watson processes, where each individual branches independently. To go beyond $b = \frac{1}{2}$ one would need a similar result as Lemma 4.4.4.

A promising approach to obtain a comparable result to Lemma 4.4.4 would be the following. Consider the random variables $(\tilde{Y}_i, i \in [N])$ such that for $1 \leq i \leq \lceil N^{b+\delta} \rceil$ we have that \tilde{Y}_i are independent copies of Y and for $i > \lceil N^{b+\delta} \rceil$ we let $\tilde{Y}_i = Y_i$. Again, we define $\tilde{\mathcal{W}} = (\tilde{W}_1, \dots, \tilde{W}_N)$

$$\tilde{W}_i := \frac{Y_i}{\sum_{i=1}^N \tilde{Y}_i} = \frac{Y_i}{\sum_{i=1}^{\lceil N^{b+\delta} \rceil} \tilde{Y}_i + \sum_{i=\lceil N^{b+\delta} \rceil+1}^N Y_i}.$$

Note that due to this construction conditionally on $\sum_{i=1}^N \tilde{Y}_i$ the random weights \tilde{W}_i , $i \leq \lceil N^{b+\delta} \rceil$ are independent. For the resulting approximating process to have the same asymptotic probability of fixation one needs that the expectations NW_1 and $N\tilde{W}_1$ are close up to an error of order $o(s_N)$. We observe for some $\alpha > b$

$$\begin{aligned} \mathbb{P}\left(|NW_1 - N\tilde{W}_1| > N^{-\alpha}\right) &= \mathbb{P}\left(\left|N \frac{Y_1}{\sum_{i=1}^N Y_i} - N \frac{Y_1}{\sum_{i=1}^N \tilde{Y}_i}\right| > N^{-\alpha}\right) \\ &\sim \mathbb{P}\left(\left|\frac{\sum_{i=1}^N Y_i - \sum_{i=1}^N \tilde{Y}_i}{NY_1}\right| > N^{-\alpha}\right) \\ &\sim \mathbb{P}\left(N^{-\frac{b+\delta}{2}} \left|\sum_{i=1}^{\lceil N^{b+\delta} \rceil} Y_i - \tilde{Y}_i\right| > N^{1-\alpha-\frac{b+\delta}{2}}\right), \end{aligned}$$

where the above probability is small by means of the central limit theorem if one has $1 - \alpha - \frac{b}{2} > 0$, which entails $b < \frac{2}{3}$. Since the random weights $\tilde{\mathcal{W}}$ are independent only when conditioned on $\sum_{i=1}^N \tilde{Y}_i$, we deal with a branching process in random environment. This lead to the project described in Section 2.3 which is concerned with the survival probabilities of a branching process in an iid random environment. In this setting we prove a similar result as for classical Galton-Watson processes.

The remaining step in the program to extend the proof via a Galton-Watson process approximation to $b < \frac{2}{3}$ would be to prove a coupling result of the Cannings frequency process with branching processes in an iid random environment in the spirit of Lemma 4.5.1. However, so far such a result has not been obtained, since the coupling becomes more involved, see Remark 2.2.6.

2.2.5 Moderate selection and heavy-tailed offspring distributions

We now complement our synopsis of Cannings models with moderate selection with an outlook on a manuscript in preparation, joint with M. Birkner, I. Dahmer and C. Pokalyuk. This concerns the fixation probability of a slightly beneficial mutant in Cannings models, where the neutral genealogy converges properly rescaled to a class of Λ -coalescents, the so-called Beta($2 - \alpha, \alpha$) coalescents.

Definition 2.2.7 (Λ -coalescent).

A Λ -coalescent is a Markov process $(X_t, t \geq 0)$ taking its values in the partitions of \mathbb{N} , such that the restriction X_t^n to $[n]$ is also a Markov process. If X_t^n consists of k blocks, $j \leq k$ of these blocks are merging into 1 at rate $\lambda_{k,j}$, with

$$\lambda_{k,j} = \int_0^1 x^{j-2} (1-x)^{k-j} \Lambda(dx),$$

for a finite measure Λ on $[0, 1]$.

Taking Λ as the Dirac measure in 0 yields the Kingman-coalescent. An important subclass of Λ -coalescents is the class of Beta-coalescents, indexed by a parameter $\alpha \in [1, 2)$. Here Λ is the Beta($2 - \alpha, \alpha$) distribution, i.e.

$$\Lambda(dx) = \beta_{2-\alpha, \alpha}(dx) := \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha} (1-x)^{\alpha-1} dx. \quad (2.13)$$

The case $\alpha = 1$ corresponds to the well known Bolthausen-Sznitman coalescent [BS98] and [Rue87] and turns out to be a boundary case for the fixation probability, see Theorem 4 and Remark 2.2.9.

As in Section 2.2.1 we consider Cannings models admitting a paintbox construction, and like in Section 2.2.4 we assume random weights \mathscr{W} of Dirichlet type. However, in contrast to the assumptions in Section 2.2.4 we assume that the random variable Y is heavy-tailed with exponent α ,

$$\mathbb{P}(Y > y) \sim \ell_\alpha y^{-\alpha}, \quad \text{as } y \rightarrow \infty, \quad (2.14)$$

for some $\alpha \in (1, 2)$ and a constant $\ell_\alpha > 0$.

Due to Theorem 1iii) by Huillet and Möhle in [HM21] the neutral genealogy converges to a $\beta_{2-\alpha, \alpha}$ -coalescent (also compare [Sch03]) after speeding up time by the factor c_N^{-1} , with

$$c_N \sim \frac{\alpha B(2-\alpha, \alpha)}{\mu^\alpha} \frac{\ell_\alpha}{N^{\alpha-1}},$$

where $\mu = \mathbb{E}[Y]$, $\alpha \in (1, 2)$ and $B(2-\alpha, \alpha)$ denotes the beta function. Similarly as before we consider the regime of moderate selection, $s_N \sim N^{-b}$ and assume

$$b < \alpha - 1, \quad \text{implying} \quad \frac{s_N}{c_N} \rightarrow \infty, \quad \text{as } N \rightarrow \infty. \quad (2.15)$$

The assumption in (2.15) entails that selection and genetic drift are of different order, hence we speak of moderate selection in the heavy-tailed case. This is in accordance with the regimes of moderate selection discussed in Section 2.2.3 and 2.2.4, where one would have $\alpha = 2$.

Theorem 4. *Under the assumptions (2.14) and (2.15), the fixation probability π_N of a single beneficial mutant obeys*

$$\pi_N \sim c_\alpha^{-\frac{1}{\alpha-1}} s_N^{\frac{1}{\alpha-1}} \quad (2.16)$$

with $c_\alpha = \frac{\ell_\alpha \Gamma(2-\alpha)}{(\alpha-1)\mu^\alpha}$.

Remark 2.2.8. *Since $s_N^{\frac{1}{\alpha-1}} = o(s_N)$, the fixation probability in the heavy-tailed case is of smaller order than in the corresponding Cannings models in the attraction of Kingman's coalescent. Note that the variance of the offspring distribution is of order $N^{2-\alpha}$ and hence a naive application of Haldane's asymptotics would imply that the fixation probability is asymptotically of order $N^{\alpha-b-2}$, which is different from the asymptotics in (2.16).*

Let us mention here two cases, where the same exponent $\frac{1}{\alpha-1}$ is observed. In so far unpublished work by M. Birkner, I. Dahmer and B. Eldon they consider a Cannings model with moderate selection and heavy-tailed offspring distribution, but selection is incorporated through a different mechanism. They show that the fixation probability obeys the asymptotics (2.16). Furthermore, they prove that the survival probability of a supercritical branching process with infinite variance and a heavy-tailed offspring distribution with exponent α , such that the offspring expectation is given as $1 + s_N$, obeys the asymptotics $s_N^{\frac{1}{\alpha-1}}$, as $s_N \rightarrow 0$, for some constant $C > 0$.

In light of the work by Slack [Sla68], the same exponent appears in the case of critical branching processes. If one considers a branching process with mean one and a heavy tailed offspring distribution ξ with parameter $\alpha \in [1, 2)$, such that $\mathbb{P}(\xi > y) \sim y^{-\alpha} L(y)$ as $y \rightarrow \infty$ for some slowly varying function L . Then by a Tauberian Theorem (see [Fel71] XIII), one obtains that the probability of survival up to time n decays like $(\frac{1}{n})^{\frac{1}{\alpha-1}} \tilde{L}(n)$, for some slowly varying function \tilde{L} .

Lastly, we give a brief outline of the proof. We proceed similarly as in the proof of Theorem 2 and apply the duality stated in Corollary 2.2.2. Again we have

$$\pi_N = \frac{\mathbb{E}[A_{eq}^\Lambda]}{N},$$

where A_{eq}^Λ denotes the Cannings ancestral selection process with Dirichlet-type weights fulfilling (2.14) in stationarity. The process A^Λ is defined by the same two-step procedure as before: Given $A_m^\Lambda = a$ we set H as a negative binomially distributed random variable with parameters a and $1 - s_N$ and then put H balls randomly into N boxes with weights \mathscr{W} and count the number of occupied boxes (compare Section 2.2.2). The difference to Section 2.2.2 arises due to the different distribution of \mathscr{W} .

Since the neutral genealogy converges to a $\beta_{(2-\alpha, \alpha)}$ -coalescent, when time is rescaled by the factor c_N^{-1} , we expect that the ancestral selection process behaves similarly to the time rescaled line counting process of a $\beta_{(2-\alpha, \alpha)}$ -coalescent in the downward movement, but additionally branching events occur per line at rate s_N . This gives rise to the heuristic that the equilibrium of A^Λ should be concentrated at that level $n = n(N)$ where upwards and downwards drift cancel, i.e.

$$ns_N \approx c_N \sum_{j=2}^n \binom{n}{j} \lambda_{n,j} (j-1). \quad (2.17)$$

The right-hand side in (2.17) corresponds to the expected drift of the time rescaled line counting process in state n .

To simplify the right-hand side of (2.17), we observe

$$\begin{aligned} \binom{n}{j} \lambda_{n,j} &= \binom{n}{j} \frac{1}{B(2-\alpha, \alpha)} \int_0^1 x^{j-1-\alpha} (1-x)^{n-j+\alpha-1} dx \\ &= \frac{\Gamma(n+1)}{\Gamma(j+1)\Gamma(n-j+1)} \frac{B(j-\alpha, n-j+\alpha)}{B(2-\alpha, \alpha)} \\ &= \frac{n}{B(2-\alpha, \alpha)} \frac{\Gamma(j-\alpha)}{\Gamma(j+1)} \frac{\Gamma(n-j+\alpha)}{\Gamma(n-j+1)}. \end{aligned}$$

Then an application of an hypergeometric identity yields for large n

$$\begin{aligned} nc_N \sum_{j=2}^n \binom{n}{j} \lambda_{n,j} (j-1) &\sim \frac{c_N}{B(2-\alpha, \alpha)} \frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)} n^\alpha \\ &\sim n^\alpha \frac{\Gamma(2-\alpha) \ell_\alpha}{(\alpha-1) \mu^\alpha} N^{1-\alpha} = c_\alpha n^\alpha N^{1-\alpha}. \end{aligned}$$

Finally, in light of (2.17) we obtain

$$\mathbb{E} [A_{eq}^\Lambda] \sim \left(\frac{1}{c_\alpha} N^{\alpha-1} s_N \right)^{\frac{1}{\alpha-1}}.$$

and hence by Corollary 2.2.2 we arrive at

$$\pi_N = \frac{\mathbb{E} [A_{eq}^\Lambda]}{N} \sim \left(\frac{1}{c_\alpha} \right)^{\frac{1}{\alpha-1}} s_N^{\frac{1}{\alpha-1}}.$$

To show rigorously that $\mathbb{E} [A_{eq}^\Lambda]$ is asymptotically given as $\left(\frac{1}{c_\alpha} N^{\alpha-1} s_N \right)^{\frac{1}{\alpha-1}}$ we apply the following reasoning, we refer to Glynn and Zeevi for further insights [GZ08]. Denote by ν the stationary distribution of A^Λ . We aim to calculate νf , with f being the identity function, read as a column vector. For a Markov chain with transition matrix P and stationary distribution ν , we have for any column vector g ,

$$\nu P = \nu \quad \text{and} \quad \nu P g = \nu g.$$

Hence, if we can find a vector g and a constant c such that

$$P g \leq g - f + c e, \tag{2.18}$$

with $e = (1, \dots, 1)^T$ we have $\nu(g - f + c e) \leq \nu g$, implying

$$\nu f \leq c.$$

The lower bound follows in the same way by choosing g and c such that $P g \geq g - f + c e$. With this approach, estimating the expectation of A_{eq}^Λ from above reduces to find a vector g such that (2.18) is fulfilled for some $c \sim \left(\frac{1}{c_\alpha} \right)^{\frac{1}{\alpha-1}} s_N^{\frac{1}{\alpha-1}}$. It turns out that for the upper bound it is sufficient to consider "linear vectors", i.e. vectors of the form $g = (g(1), \dots, g(N))$ with $g(x) = a_N x$ and some $a_N \in \mathbb{R}$. This corresponds to an analysis of the expected drift of \mathcal{A}^λ in every state $x \in [N]$. For the lower bound additionally second moments of \mathcal{A}^λ have to be estimated.

Remark 2.2.9. a) *The approach described in this section might also be useful in order to close the gap $b = \frac{1}{2}$ in the case of Cannings models within the attraction of Kingman's coalescent. However, it seems necessary to control even higher moments of W_1 than under Assumption A2, in fact*

$$\mathbb{E}[W_1^{n+1}] \leq \frac{h_N}{N} \mathbb{E}[W_1^n],$$

for some $h_N \rightarrow \infty$ as $N \rightarrow \infty$ and $n \leq N/2$.

b) *An interesting boundary case are Cannings models within the attraction of the Bolthausen-Sznitman coalescent, hence $\alpha = 1$ in (2.13). Here the pair-coalescent probability c_N is of order $\log(N)^{-1}$, hence the case of moderate selection corresponds to selection of strength $s_N \sim \log(N)^{-b}$ for $0 < b < 1$, since this implies $\frac{s_N}{c_N} \rightarrow \infty$. In this case we believe that with similar techniques as in the case $\alpha \in (1, 2)$ one shows that the survival probability π_N of a single beneficial mutant obeys*

$$\pi_N \sim \exp\left((\log N)^{1-b}\right),$$

c) *At the moment we make the rather strict assumption*

$$\mathbb{P}(Y > y) \sim \ell_\alpha y^{-\alpha}, \quad \text{as } y \rightarrow \infty,$$

for some constant $\ell_\alpha > 0$. It should be possible to consider general Pareto-distributions, i.e. to assume that ℓ_α can be replaced by a slowly varying function $\ell_\alpha(y)$ at infinity. Furthermore, as in Sections 2.2.3 and 2.2.4 it should be achievable to relax the assumption $s_N \sim N^{-b}$ to the notion

$$N^{\alpha-1-\eta} < s_N < N^{-\eta}$$

for some $\eta > 0$.

2.3 Haldane's asymptotics for branching processes in an iid random environment

In this section we summarize the results and techniques of proof used in [BK21], which is Chapter 5 in this thesis. These results concern Haldane's asymptotics for supercritical branching processes in an iid random environment. We prove that Haldane's asymptotics remains valid for branching processes in an iid random environment, provided that the fluctuations induced by the environment are not too large compared to the expected drift. For the precise statement see Theorem 5 below. In the proof we make use of a connection to perpetuities, which seems not to have been described in the literature before, see however [Als21] which appeared in arXiv only a few days after [BK21].

First, we give a precise definition of a branching process in random environment and develop the necessary notation. For an introduction and further background to branching processes in random or varying environment we refer to [KV17] and [VZ93].

Denote by $\mathcal{P}(\mathbb{N}_0)$ the space of all probability measures on \mathbb{N}_0 . It turns out that it is convenient to identify a measure f with its generating function $f(t), t \in [0, 1]$

$$f(t) = \sum_{k=0}^{\infty} t^k f[k], \quad t \in [0, 1],$$

where $f[k]$ denotes the mass of f in k . Recall that the first and second factorial moment can be expressed through the generating function as

$$f'(1) = \sum_{z=1}^{\infty} z f[z], \quad f''(1) = \sum_{z=2}^{\infty} z(z-1) f[z]. \quad (2.19)$$

Endow $\mathcal{P}(\mathbb{N}_0)$ with the total variation metric and the induced Borel- σ -algebra. Hence, $\mathcal{P}(\mathbb{N}_0)$ is a measurable space, which allows to consider a random measure F on $\mathcal{P}(\mathbb{N}_0)$. A sequence $V = (F_1, F_2, \dots)$ of random measures on $\mathcal{P}(\mathbb{N}_0)$ will be called a *random environment*. Furthermore, we call V an iid random environment, if V consists of independent and identically distributed $(F_i, i \geq 1)$, thus the F_i are independent copies of a random measure F .

Definition 2.3.1 (Branching process in random environment).

Let $V = (F_1, F_2, \dots)$ be a random environment. We say that $(Z_n, n \geq 0)$ is a branching process in random environment, if Z_n obeys the recursion

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{i,n}, \quad Z_0 = 1,$$

where $(\xi_{i,n}, i \geq 1, n \geq 1)$ are conditionally on V independent and for each $n \in \mathbb{N}$ the family $(\xi_{i,n}, i \geq 1)$ is conditionally on V independent and identically distributed with distribution F_n . Thus $\xi_{i,n}$ is a copy of some random variable ξ .

If we choose $F_i = f$ almost surely for some $f \in \mathcal{P}(\mathbb{N}_0)$, we obtain a classical Galton-Watson process with offspring distribution f .

Similar as in (2.19) the (now random) first and second factorial moment can be expressed as

$$F'(1) = \sum_{z=1}^{\infty} zF[z], \quad F''(1) = \sum_{z=2}^{\infty} z(z-1)F[z].$$

In the following we consider a sequence $(Z^{(N)}, N \geq 1)$ of slightly supercritical branching processes in an iid random environment and aim to study the asymptotic behaviour of the survival probability π_N . We indicate the dependency of N by a subscript of the probabilities or expectations, for example by writing

$$\pi_N = \mathbb{P}_N(Z_{\infty} > 0) := \mathbb{P}(Z_{\infty}^{(N)} > 0)$$

for the survival probability.

Let

$$\varepsilon_N := \mathbb{E}_N[F'(1)] - 1, \quad \nu_N := \text{Var}_N(F'(1)),$$

hence ε_N denotes the excess above 1 of $\mathbb{E}[\xi]$ and ν_N denotes the variance in offspring expectation between the generations. Since we are interested in the survival probability of a slightly supercritical branching processes we assume $\varepsilon_N > 0$ and consider the setting $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$. Furthermore, we assume stabilization of the variance of ξ , as well as uniform boundedness of the 4-th moments of ξ . Specifically, we require

$$\text{Var}_N(\xi) = \sigma^2 + o(1), \quad \mathbb{E}_N[\xi^4] = O(1), \quad (2.20)$$

with $\sigma^2 > 0$. Finally, for the first moment of F we assume

$$\mathbb{E}_N[F'(1)^{-4-\delta}] = O(1), \quad \mathbb{E}_N[|F'(1) - \mathbb{E}_N[F'(1)]|^{4+\delta}] = O(\nu_N^{2+\frac{\delta}{2}}) \quad (2.21)$$

for some $\delta > 0$. The latter assumption in (2.21) allows us to control the fluctuations of $F'(1)$ sufficiently well, whereas the first assumption is more of technical importance.

Theorem 5 (Survival probability for branching processes in random environment).

Assume that $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$. Then under the Assumptions (2.20) and (2.21), we have:

a) If $\nu_N = o(\varepsilon_N)$, then the survival probability obeys

$$\pi_N \sim \frac{2\varepsilon_N}{\sigma^2} \quad \text{as } N \rightarrow \infty.$$

b) If $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$, with $0 < \rho < 2$, then

$$\pi_N \sim \frac{(2-\rho)\varepsilon_N}{\sigma^2} \quad \text{as } N \rightarrow \infty.$$

c) If $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ with $2 < \rho < \infty$ then for large N

$$\pi_N = 0$$

Remark 2.3.2. *Let us briefly comment on the results and different regimes of Theorem 5.*

- a) *The case $\nu_N = o(\varepsilon_N)$ corresponds to the situation that the variance of the fluctuations in the random mean are asymptotically of smaller order than ε_N , which is the deterministic growth rate. In this case the survival probability of branching processes in random environment obeys the same asymptotic as classical branching processes, see [Ath92] for the latter. However, note that the fluctuations of $F'(1)$ up to order $\sqrt{\nu_N}$ are possible, thus of considerably larger magnitude than ε_N .*

In light of the fact, that the accompanying Kolmogorov asymptotics for critical processes does not carry over to the case of random environments, see [GK00], it is remarkable that Haldane's asymptotics holds true for such a large regime. For critical Galton-Watson processes the probability of survival up to generation n decays like $\frac{2}{\sigma^2 n}$ [Kol38], whereas for critical branching processes in random environment the probability of survival up to generation n decays like $\beta n^{-\frac{1}{2}}$ for some constant $\beta > 0$, hence one has two different asymptotics. In this sense supercritical branching processes in random environment inherit more properties of classical Galton-Watson processes than their critical counterparts.

- b) *The intermediate case $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ with $\rho \in (0, 2)$ interpolates between these two boundary cases a) and c) of Theorem 5. The resulting formula can be seen as a smooth linear interpolation between the two asymptotics of a) and c) and connects the subcritical/critical regime with the supercritical one. Only in this regime the random environment influences the survival probability considerably.*
- c) *On the other side of the spectrum lies case c) of Theorem 5, where the variance of the random mean is of the same order as the expected deterministic growth and so large that in fact the process is no longer supercritical. Indeed by checking the classical condition for subcriticality, which is $\mathbb{E}_N[\log F'(1)] < 0$ ([Smi68] and [SW69]), we see that the process in this regime is subcritical for N big enough, therefore extinction is almost certain.*

For the boundary case $\rho = 2$ between subcritical and supercritical behaviour, a finer analysis would be necessary. Investigating the proof of Theorem 5 c) in Chapter 5, it is apparent that higher moments of $F'(1)$ determine whether the process is sub- or supercritical or even critical.

The proof of Theorem 5 relies on the subsequent proposition, which gives an expression for the survival probability π for any supercritical branching process in an iid random environment, under conditions which are only slightly stronger than the usual conditions for supercriticality [SW69], i.e. $\mathbb{E}[\log F'(1)] > 0$. A similar approach has been used to determine the decay of the survival probability in the subcritical case [GK00]. Note that the following approach does not use the familiar fixed point characterization of π .

For any probability generating function F with positive finite mean $F'(1) > 0$ we introduce the *shape function* $\varphi(t)$ via the equation

$$\frac{1}{1 - F(t)} = \frac{1}{F'(1)(1 - t)} + \varphi(t), \quad t \in [0, 1).$$

By means of a Taylor expansion $\varphi(t)$ can be continuously extended to the entire unit interval by setting

$$\varphi(1) = \frac{F''(1)}{2F'(1)^2}.$$

In the following we consider for all $k \in \mathbb{N}$ the random shape function corresponding to the environment F_k in generation k . We denote this function by $\varphi_k(t)$, $t \in [0, 1]$, and abbreviate

$$\mu_k = \prod_{i=1}^k F'_i(1), \quad \mu_0 = 1.$$

Proposition 2.3.3. *Let $(Z_n, n \geq 0)$ be a branching process in an iid random environment $V = (F_1, F_2, \dots)$ consisting of independent copies of F . Assume*

$$0 < \mathbb{E} [\log F'(1)] < \infty \text{ and } \mathbb{E} [\log^+ F''(1)] < \infty.$$

Then the survival probability π can be expressed as

$$\pi = \mathbb{E} \left[\frac{1}{X} \right],$$

with

$$X := \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(\mathbb{P}(Z_{\infty} = 0 \mid Z_{k+1} = 1, V))}{\mu_k} < \infty \quad a.s.$$

Note that the assumptions in Proposition 2.3.3 are only slightly stronger than the classical requirements for supercriticality, i.e. $\mathbb{E} [\log F'(1)] > 0$. Further, the result holds equally true if one considers a branching process in a stationary ergodic random environment.

With Proposition 2.3.3 to prove Theorem 5 it remains to determine the asymptotic behaviour of $\mathbb{E} \left[\frac{1}{X} \right]$ under the assumptions of Theorem 5. Since we are in the slightly supercritical setting it seems reasonable to expect that the probability of extinction should converge to 1 as $N \rightarrow \infty$, hence $\mathbb{P}_N(Z_{\infty} = 0 \mid Z_{k+1} = 1, V) \rightarrow 1$ as $N \rightarrow \infty$. Therefore, we consider the approximating random variable Y of X , defined via

$$Y := \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(1)}{\mu_k},$$

which is shown to be a reasonable approximation, in the sense that $\varepsilon_N(X - Y) \rightarrow 0$ in probability, in Lemma 5.4.5.

Observe that Y fulfils the stochastic recursion

$$Y = \varphi_1(1) + \sum_{k=1}^{\infty} \frac{\varphi_{k+1}(1)}{\mu_k} \stackrel{d}{=} \varphi(1) + \frac{1}{F'(1)} Y, \quad (2.22)$$

with $\varphi(1)$ and $F'(1)$ being independent of Y . The recursion (2.22) is sometimes called *annuity equation* in a financial context [BDM16] and Y is called a *perpetuity*.

For random variables Y fulfilling (2.22) with nonnegative (A, B) such that

$$Y \stackrel{d}{=} A + BY,$$

with

$$\beta_N := 1 - \mathbb{E}_N[B], \quad \gamma_N := \text{Var}_N(B),$$

we prove the following limiting theorem for perpetuities with vanishing interest rates. This theorem might be useful in other contexts than the one presented here. In this context we also mention the recent work by Alsmeyer [Als21] on branching processes in random environment and perpetuities.

Theorem 6 (Limiting theorem for perpetuities).

Assume that in the limit $N \rightarrow \infty$, we have

$$\beta_N \rightarrow 0, \quad \gamma_N \rightarrow 0, \quad \mathbb{E}_N[A] = \alpha + o(1), \quad \mathbb{E}_N[A^{1+\delta}] = O(1),$$

with $\alpha > 0$ and some $\delta > 0$, furthermore $\mathbb{E}_N[|B - 1|^{2+\delta}] = o(\beta_N + \gamma_N)$. Additionally assume

$$\frac{\beta_N}{\gamma_N} \rightarrow \hat{\rho} \quad \text{with} \quad -\frac{1}{2} < \hat{\rho} \leq \infty.$$

Then we have:

- i) If $\hat{\rho} = \infty$, then $\beta_N Y$ converges to α in probability.
- ii) If $\hat{\rho} \in (-1/2, \infty)$, then $\gamma_N Y$ is asymptotically inverse-gamma distributed, with density $\frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x} dx$ on \mathbb{R}^+ and parameters $(a, b) = (2\hat{\rho} + 1, 2\alpha)$.

The proof of Theorem 6 uses classical tools, convergence of the Laplace-transforms to the corresponding limiting Laplace-transforms is established with the help of Lemma 5.3.1 and tightness of the sequence $(\beta_N + \gamma_N)Y$ is shown in Lemma 5.3.2.

The proof of Theorem 5 then follows with the help of some technical results immediately. By an application of Theorem 6 one obtains that $\varepsilon_N Y$ converges in probability to $\frac{\sigma^2}{2}$, if $\frac{\nu_N}{\varepsilon_N} \rightarrow 0$ as $N \rightarrow \infty$. Whereas if $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ with $0 < \rho < 2$, then $\nu_N Y$ is asymptotically inverse gamma distributed with parameters $(a, b) = \left(\frac{2-\rho}{\rho}, \sigma^2\right)$.

The proof of Theorem 5 a) and b) is then concluded via the asymptotics

$$\pi_N = \mathbb{E} \left[\frac{1}{X} \right] \sim \mathbb{E} \left[\frac{1}{Y} \right] \sim \begin{cases} \frac{2\varepsilon_N}{\sigma^2} & \text{if } \nu_N/\varepsilon_N \rightarrow 0 \\ \frac{2\varepsilon_N}{\sigma^2} \left(1 - \frac{\rho}{2}\right) & \text{if } \nu_N/\varepsilon_N \rightarrow \hat{\rho} \in (0, 2), \end{cases}$$

which are shown with the help of Lemma 5.4.4 and Lemma 5.4.5.

Chapter 3

Haldane's formula in Cannings models: The case of moderately weak selection

We introduce a Cannings model with directional selection via a paintbox construction and establish a strong duality with the line counting process of a new *Cannings ancestral selection graph* in discrete time. This duality also yields a formula for the fixation probability of the beneficial type. Haldane's formula states that for a single selectively advantageous individual in a population of haploid individuals of size N the probability of fixation is asymptotically (as $N \rightarrow \infty$) equal to the selective advantage s_N divided by half of the offspring variance. For a class of offspring distributions within Kingman attraction we prove this asymptotics for sequences s_N obeying $N^{-1} \ll s_N \ll N^{-1/2}$, which is a regime of "moderately weak selection". It turns out that for $s_N \ll N^{-2/3}$ the Cannings ancestral selection graph is so close to the ancestral selection graph of a Moran model that a suitable coupling argument allows to play the problem back asymptotically to the fixation probability in the Moran model, which can be computed explicitly.

3.1 Introduction

In population genetics the standard model for the neutral reproduction of haploid individuals in discrete time and with constant population size N is the classical *Wright-Fisher model*: the offspring numbers from one generation to the next arise by throwing N times a die with N faces and are thus $\text{Multinomial}(N; 1/N, \dots, 1/N)$ -distributed. This is a special instance within the general class of *Cannings models*, see [Can74] and [Ewe04, Chapter 3.3], where the offspring numbers are assumed to be exchangeable and to sum to N .

In the Wright-Fisher model *directional selection* can be included by appropriately biasing the weights of individuals that have a selectively beneficial type or a wildtype. While for general Cannings models it is not completely clear how to incorporate selection, a biasing of weights can be done in a natural way for the large class of Cannings models that admit a *paintbox representation* (in the sense that the N -tuple of offspring numbers is *mixed multinomial*, see Section 3.2.1).

For this class of models we introduce a graphical representation which extends to the case with directional selection, and leads to a time discrete version of the ancestral selection graph that was developed by Krone and Neuhauser in [KN97] for the (continuous time) Moran model.

Recently González Casanova and Spanò constructed in [GS18] an ancestral selection graph for a special class of Cannings models. While their construction relies on analytic arguments, we provide here a probabilistic construction which works for a wider class of models and also gives a clear interpretation of the role of the geometric distribution of the number of potential parents in this context. This construction will be explained in Section 3.5. We will prove a sampling duality between the Cannings frequency process and the line counting process of the discrete ASG (alias *Cannings ancestral selection process* or *CASP*), see Theorem 7. This also allows to obtain a handsome representation of the fixation probability of the beneficial type in terms of the expected value of the CASP in equilibrium, see Corollary 3.3.2.

The calculation of fixation probabilities is a prominent task in mathematical population genetics; for a review including a historical overview see [PW08]. A classical idea going back to Haldane, Fisher and Wright (see [Hal27], [Fis23] and [Wri31]) and known as *Haldane's formula*, is to approximate the probability of fixation of a beneficial allele with small selective advantage s by the survival probability $\pi(s)$ of a supercritical, near-critical Galton-Watson branching process,

$$\pi(s) \sim \frac{s}{\rho^2/2} \text{ as } s \rightarrow 0 \quad (3.1)$$

where ρ^2 is the offspring variance and $1 + s$ is the expected offspring size.

In Remark 3.3.4b) we will briefly discuss perspectives and frontiers of a derivation of (3.1) in terms of a branching process approximation. Couplings with Galton-Watson processes were used by González Casanova et al. in [GKWY17] to prove that (3.1) indeed gives the asymptotics for the fixation probability for a class of Cannings models (with mixed multi-hypergeometric offspring numbers) that arise in the context of experimental evolution. This was achieved under the assumption that $s_N \sim N^{-b}$ with $0 < b < 1/2$, i.e. for a moderately strong selection. There, the question remained open if (3.1) also captures the asymptotics of the fixation probability for $s_N \sim N^{-b}$ with $1/2 \leq b < 1$.

For the case $1/2 < b < 1$, our second body of main results gives an affirmative answer for subclasses of Cannings models admitting a paintbox representation, and in particular also for the Wright-Fisher model with selection. In Theorem 8a) we prove *Haldane's formula* under the condition

$$N^{-1+\eta} \leq s_N \leq N^{-2/3-\eta} \quad (3.2)$$

(with $\eta > 0$) for a class of paintboxes that satisfy in particular Möhle's condition (which guarantees that the coalescents of the neutral Cannings model are in the domain of attraction of Kingman's coalescent, see [Möh00]). Under these assumptions we show in Section 3.6 that the CASP is close to the ASG line counting process of a corresponding Moran model over a long period of time. Indeed, for a Moran model with directional selection, a representation of the fixation probability in terms of the ASG line counting process is valid, and the fixation probability can be calculated explicitly; this we explain in Section 3.4.

Relaxing (3.2), in Theorem 8b) we prove for sequences (s_N) with

$$N^{-1+\eta} \leq s_N \leq N^{-1/2-\eta} \quad (3.3)$$

under more restrictive moment conditions on the paintbox. Examples fulfilling these moment conditions include the Wright-Fisher case as well as paintboxes that are of Dirichlet-type, for more details see Section 3.3.2. The main tool of the proof under

these conditions is a concentration result on the equilibrium distribution of the CASP, see Section 7. This yields a sufficiently good estimate of the expected value of the CASP in equilibrium to show Haldane’s formula by means of the above-mentioned Corollary 3.3.2.

3.2 Cannings models with selection

3.2.1 A paintbox representation for the neutral reproduction

In a neutral Cannings model with population size N , the central concept is the exchangeable N -tuple $\nu = (\nu_1, \dots, \nu_N)$ of *offspring sizes*, with non-negative integer-valued components summing to N . A reasonably large class of such random variables ν admits a *paintbox construction*, i.e. has a mixed multinomial distribution with parameters N and \mathscr{W} , where $\mathscr{W} = (W_1, W_2, \dots, W_N)$ is an exchangeable random N -tuple of probability weights taking its values in

$$\Delta_N = \left\{ (x_1, x_2, \dots, x_N) : x_i \geq 0, \sum_{i=1}^N x_i = 1 \right\}.$$

While this is clearly reminiscent of Kingman’s paintbox representation of exchangeable partitions of \mathbb{N} , here we are dealing with a finite N . As such, obviously, not all exchangeable offspring sizes are mixed multinomial – consider e.g. a uniform permutation of the vector $(2, \dots, 2, 0, \dots, 0)$. On the other hand, the exchangeable mixed multinomials cover a wide range of applications; e.g., they can be seen as approximations of the offspring sizes in a model of experimental evolution, where at the end of each reproduction cycle N individuals are sampled *without replacement* from a union of N families with large i.i.d. sizes; see [GKQY17] and [BGPW19], where the distribution of the family sizes was assumed to be geometric with expectation $\gamma = 100$. This leads to a mixed multi-hypergeometric offspring distribution, whose analogue for $\gamma = \infty$ would be a mixed multinomial offspring distribution with $\mathcal{L}(\mathscr{W})$ the Dirichlet(1, …, 1)-distribution on Δ_N .

Let us now briefly review the graph of genealogical relationships in a Cannings model. In each generation g , the individuals are numbered by $i \in [N] := \{1, \dots, N\}$ and denoted by (i, g) . A *parental relation* between individuals in generation g and $g-1$ is defined in the following way. Let $\mathscr{W}^{(g)}, g \in \mathbb{Z}$, be i.i.d. copies of \mathscr{W} . Every individual (j, g) is assigned a parent $(V_{(j,g)}, g-1)$ in generation $g-1$ by means of an $[N]$ -valued random variable $V_{(j,g)}$ with conditional distribution $\mathbb{P}(V_{(j,g)} = i | \mathscr{W}^{(g-1)}) = W_i^{(g-1)}$, $i \in [N]$. For each $g \in \mathbb{Z}$, the random variables $V_{(j,g)}, j = 1, \dots, N$, are assumed to be independent given $\mathscr{W}^{(g-1)}$. Also, for each $g \in \mathbb{Z}$, due to the exchangeability of $(W_1^{(g-1)}, \dots, W_N^{(g-1)})$, the random variables $V_{(1,g)}, \dots, V_{(N,g)}$ are uniformly distributed on $[N]$, and in general are correlated. With this construction within one generation step we produce an exchangeable N -tuple of *offspring sizes*, i.e. the number of children for each individual $(i, g-1), i \in [N]$. Due to the assumed independence of the random variables $\mathscr{W}^{(g)}, g \in \mathbb{Z}$, the offspring sizes as well as the “assignments to parents” $(V_{(1,g)}, \dots, V_{(N,g)})$ are independent along the generations g .

3.2.2 A paintbox representation incorporating selection

We now build directional selection with strength $s_N \in (0, 1)$ into the model. Assume that each individual has one of two types, either the *beneficial type* or the *wildtype*.

Let the chances to be chosen as a parent be modified by decreasing the weight of each wildtype individual by the factor $1 - s_N$. In other words, if individual (i, g) has the wildtype the weight reduces to $\widetilde{W}_i := (1 - s_N)W_i$ and if the individual has the beneficial type the weight remains $\widetilde{W}_i := W_i$. Let $\widetilde{\mathcal{W}}^{(g)} := (\widetilde{W}_1^{(g)}, \dots, \widetilde{W}_N^{(g)})$. Given the type configuration in generation $g - 1$, the parental relations are now generated in a two-step manner: First, assign the random weights $\widetilde{\mathcal{W}}^{(g-1)}$ to the individuals in generation $g - 1$, then follow the rule

$$\mathbb{P}((i, g - 1) \text{ is parent of } (j, g) \mid \widetilde{\mathcal{W}}^{(g-1)}) = \frac{\widetilde{W}_i^{(g-1)}}{\sum_{\ell=1}^N \widetilde{W}_\ell^{(g-1)}}. \quad (3.4)$$

Individual (j, g) then inherits the type from its parent. Note that $\widetilde{\mathcal{W}}^{(g-1)}$ is measurable with respect to $\mathcal{W}^{(g-1)}$ together with the type configuration in generation $g - 1$. Because of the assumed exchangeability of the $W_i^{(g-1)}$, $i = 1, \dots, N$, the distribution of the type configuration in generation g only depends on the number of individuals in generation $g - 1$ that carry the beneficial type. Thus, formula (3.4) defines a Markovian dynamics for the type frequencies. We will denote the number of wildtype individuals in generation g by K_g , and will call $(K_g)_{g=0,1,\dots}$ a *Cannings frequency process* with parameters N , $\mathcal{L}(\mathcal{W})$ and s_N . In particular, (3.4) implies that given $\{K_{g-1} = k\}$, K_g is mixed Binomial with parameters N and $P(k, \mathcal{W})$, where

$$P(k, \mathcal{W}) = \frac{(1 - s_N) \sum_{i=1}^k W_i}{(1 - s_N) \sum_{\ell=1}^k W_\ell + \sum_{\ell=k+1}^N W_\ell}. \quad (3.5)$$

3.2.3 The Cannings ancestral selection process

Again let $N \in \mathbb{N}$, \mathcal{W} as in Section 3.2.1, and $s_N \in (0, 1)$. The *Cannings ancestral selection process (CASP)* $(A_m)_{m=0,1,\dots}$ with parameters N , $\mathcal{L}(\mathcal{W})$ and s_N counts the number of potential ancestors in generation $g - m$ of a sample taken in generation g . We will give a graphical representation in Section 3.5; in the present section we define the CASP as an $[N]$ -valued Markov chain whose one-step transition is composed of a branching and a coalescence step as follows:

Given $A_m = a$, the branching step takes a into a sum $H = \sum_{\ell=1}^a G^{(\ell)}$ of independent $\text{Geom}(1 - s_N)$ -random variables; in other words, the random variable H has a negative binomial distribution with parameters a and $1 - s_N$, and thus takes its values in $\{a, a + 1, \dots\}$. (Here and below, we understand a $\text{Geom}(p)$ -distributed random variable as describing the number of trials (and not only failures) up to and including the first success in a coin tossing with success probability p .)

The coalescence step arises (in distribution) through a two-stage experiment: first choose a random \mathcal{W} according to the prescribed distribution $\mathcal{L}(\mathcal{W})$, then, given \mathcal{W} and the number H from the branching step, place H balls independently into N boxes, where W_i is the probability that the first (second, . . . , H -th) ball is placed into the i -th box, $i = 1, \dots, N$. The random variable A_{m+1} is distributed as the number of non-empty boxes.

3.3 Main results

3.3.1 Duality of Cannings frequency and ancestral selection process

For $N \in \mathbb{N}$, \mathscr{W} as in Section 3.2.1, and $s_N \in (0, 1)$, let $(K_g)_{g \geq 0}$ be the Cannings frequency process with parameters N , $\mathcal{L}(\mathscr{W})$ and s_N as defined in Section 3.2.2, and let $(A_m)_{m \geq 0}$ be the Cannings ancestral selection process with parameters N , $\mathcal{L}(\mathscr{W})$ and s_N as defined in Section 3.2.3.

Theorem 7 (Sampling duality). *Let $g \geq 0$, and $k, n \in [N]$. Let \mathcal{J}_n be uniformly chosen from all subsets of $[N]$ of size n , and given $A_g = a$, $a \in [N]$, let \mathcal{A}_g be uniformly chosen from all subsets of $[N]$ of size a . Then we have the following duality relation*

$$\mathbb{P}(\mathcal{J}_n \subset [K_g] \mid K_0 = k) = \mathbb{P}(\mathcal{A}_g \subset [k] \mid A_0 = n). \quad (3.6)$$

Remark 3.3.1. *A strong (pathwise) version of the duality relation (3.6) will be provided by formula (3.23) in Section 3.5, which roughly spoken says that “A sample from generation g is entirely of wildtype if and only if all of its potential ancestors in generation 0 are of wildtype”.*

Expressed in terms of K_g and A_g , the sampling duality relation (3.6) becomes

$$\begin{aligned} & \mathbb{E} \left[\frac{K_g(K_g - 1) \cdots (K_g - n + 1)}{N(N - 1) \cdots (N - n + 1)} \mid K_0 = k \right] \\ &= \mathbb{E} \left[\frac{k(k - 1) \cdots (k - A_g + 1)}{N(N - 1) \cdots (N - A_g + 1)} \mid A_0 = n \right]. \end{aligned} \quad (3.7)$$

Specializing (3.7) to $k = N - 1$ and $n = N$ gives

$$\mathbb{P}(K_g = N \mid K_0 = N - 1) = 1 - \mathbb{E} \left[\frac{A_g}{N} \mid A_0 = N \right] \quad (3.8)$$

Taking the limit $g \rightarrow \infty$ in (3.8) leads to

Corollary 3.3.2. *Let A_{eq} have the stationary distribution of the Cannings ancestral selection process $(A_m)_{m \geq 0}$. The fixation probability of a single beneficial mutant is*

$$\pi_N := \lim_{g \rightarrow \infty} \mathbb{P}[K_g = 0 \mid K_0 = N - 1] = \mathbb{E} \left[\frac{A_{\text{eq}}}{N} \right]. \quad (3.9)$$

Remark 3.3.3. *In the light of Remark 3.3.1, the representation (3.9) can be interpreted as follows: With a single beneficial mutant in generation 0, the beneficial type goes to fixation if and only if the beneficial mutant is among the potential ancestors in generation 0 of the population at a late generation g . In the limit $g \rightarrow \infty$ the number of these potential ancestors is distributed as A_{eq} , and given A_{eq} , the probability that the beneficial mutant is among them is $\frac{A_{\text{eq}}}{N}$.*

3.3.2 Haldane’s formula for Cannings models with selection

Let (s_N) be a sequence in $(0, 1)$ that satisfies (3.2). For each N let

$\mathscr{W}^{(N)} = (W_1^{(N)}, \dots, W_N^{(N)})$ be as in Section 3.2.1, and assume that for some $\rho^2 \geq 1$

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^2 \right] = \frac{\rho^2}{N^2} + O(N^{-3}), \quad (3.10)$$

with the $O(\cdot)$ -terms referring to $N \rightarrow \infty$. (The requirement $\rho^2 \geq 1$ is natural because $\mathbb{E} \left[W_1^{(N)} \right] = \frac{1}{N}$.) We consider the following condition:

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^3 \right] = O(N^{-3}) \quad (3.11)$$

Note that (3.10) and (3.11) imply

$$N \mathbb{E} \left[\left(W_1^{(N)} \right)^2 \right] \rightarrow 0 \quad \text{and} \quad \mathbb{E} \left[\left(W_1^{(N)} \right)^3 \right] = o \left(\mathbb{E} \left[\left(W_1^{(N)} \right)^2 \right] \right). \quad (3.12)$$

For Cannings processes admitting a paintbox representation, (3.12) is equivalent to Möhle's condition, see [Möh00], which, in turn, is equivalent to the neutral Cannings coalescent being in the domain of attraction of a Kingman coalescent as $N \rightarrow \infty$.

We also consider the following condition which is stronger than (3.11): There exists a sequence (h_N) of natural number with the properties

$$h_N \rightarrow \infty \quad \text{and} \quad h_N = o(\ln N) \quad \text{as } N \rightarrow \infty \quad (3.13)$$

such that for all sufficiently large N and all $n \leq 2h_N$

$$\frac{N}{h_N} \mathbb{E} \left[\left(W_1^{(N)} \right)^{n+1} \right] \leq \mathbb{E} \left[\left(W_1^{(N)} \right)^n \right]. \quad (3.14)$$

As we will prove at the very end of Section 3.7, (3.14) is satisfied for all sequences of symmetric Dirichlet distributions whose sequence of parameters α_N converges as $N \rightarrow \infty$.

Theorem 8 (Haldane's formula).

Consider a sequence of Cannings frequency processes with parameters N , $\mathcal{L}(\mathcal{W}^{(N)})$ and s_N . Assume that the selection is moderately weak in the sense that (s_N) satisfies (3.3), and assume that $\mathcal{L}(\mathcal{W}^{(N)})$ satisfies (3.11) and (3.10) with $\rho \geq 1$. Then the fixation probabilities π_N of single beneficial mutants follow the asymptotics

$$\pi_N = \frac{2s_N}{\rho^2} + o(s_N) \quad \text{as } N \rightarrow \infty \quad (3.15)$$

provided one of the following additional requirements a) or b) is satisfied:

- a) (s_N) satisfies (3.2),
- b) the moments of $(W_1^{(N)})$ obey (3.14), with a sequence (h_N) that satisfies (3.13).

In particular, (3.10) and the requirement b) are satisfied in the Wright-Fisher case $W_i^{(N)} \equiv \frac{1}{N}$ (with $\rho^2 = 1$) as well as in the case where $\mathcal{L}(\mathcal{W}^{(N)})$ is a Dirichlet($\alpha_N, \dots, \alpha_N$)-distribution with $\frac{\alpha_N+1}{\alpha_N} = \rho^2 + O(1/N)$ as $N \rightarrow \infty$.

The proof of part a) will be given in Section 3.6 and that of part b) in Section 3.7. Here we give brief sketches of the proofs.

Part a): We know that the asymptotics (3.15) holds for the fixation probabilities π_N^M (starting from a single beneficial mutant) in a sequence of Moran(N)-models with neutral reproduction rate $\rho^2/2$ (or equivalently with pair coalescence probability ρ^2)

and selection strength s_N . Indeed, in Section 3.4 we will argue that

$$\pi_N^M = \frac{\mathbb{E} \left[B_{\text{eq}}^{(N)} \right]}{N}, \quad (3.16)$$

where $B_{\text{eq}}^{(N)}$ has the stationary distribution of the line counting process $(B_r^{(N)})_{r \geq 0}$ in the ancestral selection graph belonging to the Moran model. As observed in [Cor17], $B_{\text{eq}}^{(N)}$ is a binomially distributed random variable with parameters N and $p_N := \frac{2s_N}{2s_N + \rho^2}$ that is conditioned not to vanish. In particular,

$$\frac{\mathbb{E} \left[B_{\text{eq}}^{(N)} \right]}{N} = p_N \frac{1}{1 - (1 - p_N)^N}, \quad (3.17)$$

which because of $s_N \rightarrow 0$ and $s_N N \rightarrow \infty$ equals $\frac{2s_N}{\rho^2} + o(s_N)$.

We show in Section 3.6 that thanks to conditions (3.10) and (3.11) and Assumption (3.2), we can couple $(A_m^{(N)})_{m \geq 0}$ and $(B_r^{(N)})_{r \geq 0}$ long enough to ensure $\mathbb{E} \left[A_{\text{eq}}^{(N)} \right] = \mathbb{E} \left[B_{\text{eq}}^{(N)} \right] (1 + o(1))$, which proves Theorem 8a).

Part b): In Section 3.7 we show that the CASP needs a polynomially long time to enter a region of size $N^{-\varepsilon} s_N N$ around the “center” $\frac{2}{\rho^2} s_N N$, for ε small enough, but does not leave this region up to any polynomially long time with sufficiently high probability. This yields that the expectation of the CASP in equilibrium is $\frac{2}{\rho^2} s_N N (1 + o(1))$ and gives with Corollary 3.3.2 Haldane’s formula.

Remark 3.3.4. *a) (Moderately strong selection and the equilibrium of the CASP) Our method of proof of Theorem 8 employs Corollary 3.3.2, for part a) by the just indicated coupling of (A_m) and (B_r) and for part b) via a concentration analysis of the equilibrium distribution of (A_m) . In both cases it is essential that s_N is not too large, which is responsible for the upward jumps by branching of (A_m) and hence the number of lines of a CASP in equilibrium is not too large; this is guaranteed by the conditions (3.2) and (3.3), respectively. In particular, the backwards approach relies on the assumption $s_N \ll N^{-1/2}$. To apply Corollary 3.3.2 we need to show that the expectation of the CASP in equilibrium is $\frac{2}{\rho^2} N s_N (1 + o(1))$. We show that the CASP needs an at most polynomially long time to enter a central region of the CASP from outside, but does not leave this central region up to any polynomially long time with sufficiently high probability. For this purpose we couple \mathcal{A} with a random walk that makes jumps only of limited size. To show that large jumps (upwards or downwards) are negligible, we make use of the assumption $s_N \ll N^{-1/2}$. The probability that the CASP (in a state close to the “center”, that is of the order of $N s_N$) makes a jump at least of size h_N (with $1 \ll h_N \ll N s_N$) upwards can essentially be estimated by the probability that for at least h_N individuals at the branching step at least 2 lines are generated. The probability that at the branching step for a given individual 2 lines are generated is $\approx s_N$. There are $(N s_N)! / (N s_N - h_N)! = \Theta((N s_N)^{h_N})$ possibilities to choose h_N individuals out of $N s_N$ many individuals. The probability that for individuals $1, \dots, h_N$ at the branching step 2 lines are generated is $s_N^{h_N}$. Consequently, the probability for a jump of at least of size h_N upwards can be estimated from above by $O(s_N^{h_N} (N s_N)^{h_N}) = O((N s_N^2)^{h_N})$, which is of negligible size, if $s_N \ll N^{-1/2}$.*

b) (Moderately weak selection and Galton-Watson approximations) A regime for which it is possible to derive (3.15) by means of a Galton-Watson approximation is that of moderately strong selection $1 \gg s_N \gg N^{-1/2}$. A proof of this assertion (under

somewhat more restrictive moment conditions on the paintbox \mathcal{W} than those in Theorem 8 is the subject of a paper in preparation; see also the discussion in the paragraph following (3.1) in the Introduction. Together with the approach of the present paper this does not yet cover the case $s_N \sim N^{-1/2}$; we conjecture that Haldane's formula is valid also for this particular exponent. Here is a quick argument which explains the relevance of the exponent $1/2$ as a border for the applicability of a Galton-Watson approximation. The beneficial type is with high probability saved from extinction if the number of individuals of the beneficial type exceeds (of the order of) s_N^{-1} . Hence, for a proof via approximations with Galton-Watson processes one wants couplings of the CASP with GW-processes to hold until this number of beneficial individuals is reached. However, a Galton-Watson approximation works only until there is an appreciable amount of "collisions" between the offspring of the beneficial individuals in a branching step, since collisions destroy independence. By well-known "birthday problem" considerations, such an amount of collisions happens as soon as there are (of the order of) $N^{1/2}$ beneficial individuals in the population. Consequently, for the GW-approach we require $s_N^{-1} \ll N^{1/2}$.

- c) (Coupling with a Moran model) The relevance of Condition (3.2) for our proof of part a) of Theorem 8 (whose details are worked out in Section 3.6) can heuristically be seen as follows. An inspection of the jump probabilities described in Section 3.2.3 shows that in a regime of negligible multiple collisions the quantity $a_{\text{eq}}^{(N)} := N \frac{s_N}{\rho^2/2}$ is an asymptotic center of attraction for the dynamics of $A^{(N)}$. Since $B^{(N)}$ is close to its equilibrium after a time interval of length $s_N^{-(1+\delta)}$ we require the coupling of $A^{(N)}$ and $B^{(N)}$ to hold for $s_N^{-(1+\delta)}$ many generations. This works if within this time interval the number of potential ancestors $A_m^{(N)}$ makes at most jumps of size 1 in each generation. The number of potential parents decreases by 1 if a single pair coalesces. Near $a_{\text{eq}}^{(N)}$ the number of pairs within $A_{\text{eq}}^{(N)}$ is of the order $(a_{\text{eq}}^{(N)})^2$. Hence, the probability of a pair coalescence per generation is of the order $(a_{\text{eq}}^{(N)})^2/N$, and the probability of more than a single pair coalescence per generation is of the order $(a_{\text{eq}}^{(N)})^4/N^2$, or equivalently, of the order $N^2 s_N^4$, since this probability can be estimated by the probability for two pair coalescences, see Lemma 3.6.1. Analogously, the probability that the number of potential parents increases by more than 1 is also of order $N^2 s_N^4$. Hence, the probability that jumps at most of size 1 occur for $s_N^{-(1+\delta)}$ generations, that have to be considered, is of the order $(1 - N^2 s_N^4)^{s_N^{-(1+\delta)}}$. This probability is high, if $N^2 s_N^{3-\delta} \ll 1$, which corresponds to the upper bound in (3.2).
- d) (Possible generalisations) The introduced duality method for Cannings models with selection may well prove beneficial also in more general settings. The construction of the Cannings ancestral selection graph given in Section 3.5 can for example also be carried out in a many-island situation, with migration between islands in discrete or continuous time. This should then lead to generalizations of Theorems 7 and 8. Under assumption (3.2), and with an appropriate scaling of the migration probabilities, one might expect that the Cannings ancestral selection graph is again close to the (now structured) Moran ancestral selection graph.

3.4 Haldane's formula in the Moran model

In a two-type Moran model with constant population size N and directional selection (see e.g. [Dur08, Chapter 6]), each individual reproduces at a constant rate $\gamma/2$, $\gamma > 0$,

and individuals of the beneficial type reproduce with an additional rate $s_N > 0$. Let Y_t^N be the number of wildtype individuals at time t , and let $(B_r^N)_{r \geq 0}$ be the counting process of potential ancestors traced back from some fixed time. The process $(B_r)_{r \geq 0}$, which we call the Moran ancestral selection process (or MASP for short), is a Markov jump process with jumps from k to $k + 1$ at rate $k s_N \frac{N-k}{N}$ for $1 \leq k \leq N - 1$ and from k to $k - 1$ at rate $\frac{\gamma}{N} \binom{k}{2}$, see [KN97]. The well-known graphical representation of the Moran model yields a strong duality between $(Y_t^N)_{t \geq 0}$ and $(B_r^N)_{r \geq 0}$. Stated in words this says that a sample $J \subset [N]$ at time t consists solely of wildtype individuals if and only if all the potential ancestors of J are wildtype. This immediately leads to the following (hypergeometric) sampling duality

$$\begin{aligned} & \mathbb{E} \left[\frac{Y_t^N (Y_t^N - 1) \cdots (Y_t^N - (n - 1))}{N(N - 1) \cdots (N - (n - 1))} \middle| Y_0^N = k \right] \\ &= \mathbb{E} \left[\frac{k(k - 1) \cdots (k - (B_t^N - 1))}{N(N - 1) \cdots (N - (B_t^N - 1))} \middle| B_0^N = n \right] \end{aligned}$$

where $t \geq 0$ and $k, n \in [N]$. Specializing the latter to $n = N$ and $k = N - 1$ we obtain as in Corollary 3.3.2 that the probability π_N^M of fixation of a single beneficial mutant is given by (3.16). Thus π_N^M is given by the r.h.s. of (3.17) with ρ^2 replaced by γ .

In particular, for $s_N = \frac{\alpha}{N}$, $\alpha > 0$ (the case of weak selection) this specializes to Kimura's formula [Kim57]

$$\pi_N^M = \frac{2s_N}{\gamma} \frac{1}{1 - e^{-\frac{2\alpha}{\gamma}}} (1 + o(N^{-1})).$$

For $N^{-\eta} \geq s_N \geq N^{-1+\eta}$ (the case of moderate selection) we obtain Haldane's formula

$$\pi_N^M = \frac{2s_N}{\gamma} (1 + o(s_N))$$

and for $s > 0$ (the case of strong selection) this results in

$$\pi_N^M = \frac{2s}{2s + \gamma} (1 + O(N^{-1})).$$

3.5 The Cannings ancestral selection graph. Proof of Theorem 7

We now define the Cannings ancestral selection graph, i.e. the graph of potential ancestors in a Cannings model with directional selection as announced in Section 3.2.3. The final harvest of this section will be the proof of Theorem 7.

While the branching-coalescing structure of the Moran ancestral selection graph and the sampling duality stated in Section 3.4 serve as a conceptual guideline, the ingredients of the graphical construction turn out to be quite different from the Moran case, not least because of the discrete generation scheme.

We first describe how, given $\mathscr{W}^{(g-1)}$ and the configuration of types of the individuals $(i, g - 1)$ in generation $g - 1$, the *parent* (as well as the type) of an individual (j, g) is constructed from a sequence of i.i.d. uniform picks from the unit square.

After this we describe how, given $\mathscr{W}^{(g-1)}$ (and without prior knowledge of the type configuration in generation $g - 1$), the just mentioned i.i.d. uniform picks from the unit square lead to the *potential parents* of an individual (j, g) . The latter form a random subset of $[N] \times \{g - 1\}$.

To this purpose, as illustrated in Figure 3.1, think of the two axes of the unit square as being partitioned in two respectively N subintervals. The two subintervals that partition the horizontal unit interval are $[0, 1 - s_N]$ and $(1 - s_N, 1]$. The N subintervals of the vertical unit interval have lengths $W_1^{(g-1)}, \dots, W_N^{(g-1)}$; we call these subintervals $\mathcal{I}_1^{(g-1)}, \dots, \mathcal{I}_N^{(g-1)}$. Let

$$\begin{aligned}\mathcal{B}^{(g-1)} &:= \{i \in [N] : (i, g-1) \text{ is of beneficial type}\}, \\ \mathcal{C}^{(g-1)} &:= \{i \in [N] : (i, g-1) \text{ is of wildtype}\},\end{aligned}$$

and define

$$\Gamma^{(g-1)} := \bigcup_{i \in \mathcal{B}^{(g-1)}} [0, 1] \times \mathcal{I}_i^{(g-1)} \cup \bigcup_{i \in \mathcal{C}^{(g-1)}} [0, 1 - s_N] \times \mathcal{I}_i^{(g-1)}. \quad (3.18)$$

Definition 3.5.1. For fixed $j \in [N]$ and $g \in \mathbb{Z}$, let $U^{(j,g,1)}, U^{(j,g,2)}, \dots$ be a sequence of independent uniform picks from $[0, 1] \times [0, 1]$ and put

$$\gamma(j, g) := \min\{\ell : U^{(j,g,\ell)} \in \Gamma^{(g-1)}\}.$$

Given $\mathcal{B}^{(g-1)}, \mathcal{C}^{(g-1)}, \mathcal{W}^{(g-1)}$ and $U^{(j,g,1)}, U^{(j,g,2)}, \dots$, there is a.s. a unique $p(j, g) \in [N]$ for which $U^{(j,g,\gamma(j,g))} \in [0, 1] \times \mathcal{I}_{p(j,g)}^{(g-1)}$. The individual $(p(j, g), g-1)$ is defined to be the parent of (j, g) .

Decreeing that individual (j, g) inherits the type of its parent, we obtain that a.s.

$$\begin{aligned}\{(j, g) \text{ is of wildtype}\} &:= \{(p(j, g), g-1) \text{ is of wildtype}\} \\ &= \{U^{(j,g,\gamma(j,g))} \in [0, 1 - s_N] \times \bigcup_{i \in \mathcal{C}^{(g-1)}} \mathcal{I}_i^{(g-1)}\}\end{aligned}$$

We thus get the transport of $\mathcal{C}^{(g-1)}$ to the next generation g by putting

$$\mathcal{C}^{(g)} := \{j \in [N] : (j, g) \text{ is of wildtype}\}.$$

Remark 3.5.2. The $[N]$ -valued process $|\mathcal{C}^{(g)}|$, $g = 0, 1, \dots$ is a Cannings frequency process with parameters N , $\mathcal{L}(\mathcal{W})$ and s_N , as defined in Section 3.2.2. Indeed, given $\mathcal{C}^{(g-1)}$ and $\mathcal{W}^{(g-1)}$, the random variables $U^{(j,g,\gamma(j,g))}$, $j = 1, \dots, N$, are independent and uniformly distributed on $\Gamma^{(g-1)}$, hence (3.19) and the exchangeability of the components of $\mathcal{W}^{(g-1)}$ implies that given $\{|\mathcal{C}^{(g-1)}| = k\}$ (and with an arbitrary allocation of these k elements in the set $[N]$), the random variable $|\mathcal{C}^{(g)}|$ has a mixed Binomial distribution with parameters N and $P(k, \mathcal{W})$ specified by (3.5).

Let us now turn to a situation in which the type configuration of the previous generation is not given, i.e. in which the sets $\mathcal{B}^{(g-1)}$ and $\mathcal{C}^{(g-1)}$ and hence also the set $\Gamma^{(g-1)}$ is not known a priori.

Definition 3.5.3. i) For fixed $j \in [N]$ and $g \in \mathbb{Z}$, let $U^{(j,g,1)}, U^{(j,g,2)}, \dots$ be as in Definition 3.5.1 and define

$$G(j, g) := \min\{\ell : U^{(j,g,\ell)} \in [0, 1 - s_N] \times [0, 1]\}. \quad (3.20)$$

We call $(i, g-1)$ a potential parent of (j, g) if $U^{(j,g,\ell)} \in [0, 1] \times \mathcal{I}_i^{(g-1)}$ for some $\ell \leq G(j, g)$. Similarly, we call $(i, g-2)$ a potential grandparent of (j, g) if $(i, g-2)$

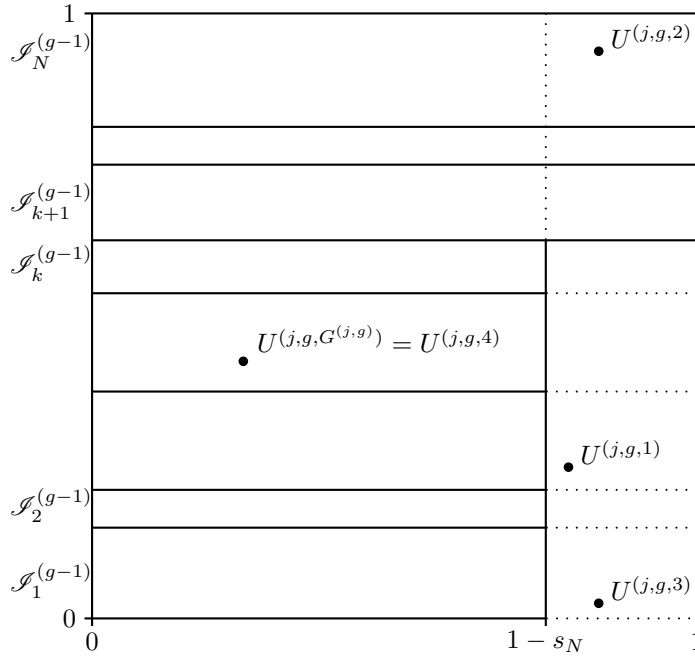


FIGURE 3.1: This figure illustrates a case in which $\mathcal{C}^{(g-1)} = \{1, \dots, k\}$, $\mathcal{B}^{(g-1)} = \{k+1, \dots, N\}$, $\gamma(j, g) = 2$, $G(j, g) = 4$. The individual (j, g) is of beneficial type. Since in this example $\gamma(j, g)$ is strictly smaller than $G(j, g)$, the individual (j, g) must be of beneficial type.

is a potential parent of a potential parent of (j, g) . By iteration this extends to the definition of the set $\mathcal{A}_m^{(j, g)}$ of potential ancestors of (j, g) in generation $g - m$, $m \geq 1$, with $\mathcal{A}_0^{(j, g)} := \{(j, g)\}$.

- ii) For a set $\mathcal{J} \subset [N]$ and for $m \geq 0$ let $\mathcal{A}_m^{(\mathcal{J}, g)} := \bigcup_{j \in \mathcal{J}} \mathcal{A}_m^{(j, g)}$ be the set of potential ancestors of $\mathcal{J} \times \{g\}$ in generation $g - m$. Moreover, let $A_m^{(\mathcal{J}, g)} := |\mathcal{A}_m^{(\mathcal{J}, g)}|$ be the number of potential ancestors of $\mathcal{J} \times \{g\}$ in generation $g - m$.

The a.s. equality of events asserted in the following lemma is both crucial and elementary.

Lemma 3.5.4. For $j, g, U^{(j, g, 1)}, U^{(j, g, 2)}, \dots, \gamma(j, g)$ as in Definition 3.5.1 and $G(j, g)$ as in (3.20),

$$\begin{aligned} \{U^{(j, g, \gamma(j, g))} \in [0, 1 - s_N] \times \bigcup_{i \in \mathcal{C}^{(g-1)}} \mathcal{I}_i^{(g-1)}\} \\ \stackrel{\text{a.s.}}{=} \{U^{(j, g, 1)}, \dots, U^{(j, g, G(j, g))} \in [0, 1] \times \bigcup_{i \in \mathcal{C}^{(g-1)}} \mathcal{I}_i^{(g-1)}\}. \end{aligned} \quad (3.21)$$

Proof. To see that the l.h.s. almost surely implies the r.h.s., consider the first pick that falls into the area $\Gamma^{(g-1)}$ and assume that it lands in a horizontal stripe belonging to a wildtype individual in generation $g - 1$. Then this must be also the first one of the picks that lands in $[0, 1 - s_N] \times [0, 1]$, and no one of the preceding picks could have landed in a horizontal stripe belonging to a beneficial individual in generation $g - 1$. Conversely, to see that the r.h.s. (3.21) a.s. implies the l.h.s., consider the first pick that lands in $[0, 1 - s_N] \times [0, 1]$. If all the picks up to and including this pick have

landed in a horizontal stripe belonging to a wildtype individual in generation $g - 1$, then also the first pick that falls into the area $\Gamma^{(g-1)}$ must land in a horizontal stripe belonging to a wildtype individual in generation $g - 1$. \square

Combining (3.19) and (3.21) with Definition 3.5.3 we see that for all $g \in \mathbb{N}$ and all $\mathcal{J} \subset [N]$

$$\{\mathcal{J} \subset \mathcal{C}^{(g)}\} \stackrel{\text{a.s.}}{=} \{\mathcal{A}_1^{(\mathcal{J},g)} \subset \mathcal{C}^{(g-1)}\} \quad (3.22)$$

Iterating (3.22) we arrive at

$$\{\mathcal{J} \subset \mathcal{C}^{(g)}\} \stackrel{\text{a.s.}}{=} \{\mathcal{A}_g^{(\mathcal{J},g)} \subset \mathcal{C}^{(0)}\}. \quad (3.23)$$

It is obvious that the random variables $G(j, g)$ defined in (3.20) are independent of the $\mathcal{W}^{(g')}$, $g' \in \mathbb{Z}$, and have the property

$$G(j, g), g \in \mathbb{Z}, j \in [N], \quad \text{are independent and } \text{Geom}(1 - s_N) \text{ distributed.} \quad (3.24)$$

This leads directly to the following observation on the number of potential ancestors.

Remark 3.5.5. *Let $g \in \mathbb{Z}$ and $\mathcal{J} \subset [N]$ be fixed.*

- i) *The process $|\mathcal{A}_m^{(\mathcal{J},g)}|$, $m = 0, 1, \dots$, is a Cannings ancestral selection process (CASP) with parameters N , $\mathcal{L}(\mathcal{W})$ and s_N , as defined in Section 3.2.3. Indeed, each transition consists of a branching and a coalescence step, where only the latter depends on the $\mathcal{W}^{(g')}$, $g' \in \mathbb{Z}$. Specifically, given $|\mathcal{A}_m^{(\mathcal{J},g)}| = a$, let H have a negative binomial distribution with parameters a and $1 - s_N$. Given $H = h$, $|\mathcal{A}_{m+1}^{(\mathcal{J},g)}|$ is distributed as the number of distinct outcomes in h trials, which given $\mathcal{W}^{(g-m-1)}$ are independent and follow the probability weights $\mathcal{W}^{(g-m-1)}$.*
- ii) *For $m \geq 1$ the exchangeability of the components of \mathcal{W} implies that, given $|\mathcal{A}_m^{(\mathcal{J},g)}| = a$, the set $\mathcal{A}_m^{(\mathcal{J},g)}$ is a uniform pick of all subsets of $[N]$ of cardinality a .*

Proof of Theorem 7. Let k, n and \mathcal{J}_n be as in Theorem 7. In (3.23) we choose $\mathcal{J} := \mathcal{J}_n$ and $\mathcal{C}_0 := [k]$. Then

$$\mathbb{P}(\mathcal{J}_n \subset [K_g] \mid K_0 = k) = \mathbb{P}(\mathcal{J}_n \subset \mathcal{C}^{(g)}) = \mathbb{P}(\mathcal{A}_g^{(\mathcal{J}_n,g)} \subset \mathcal{C}^{(0)}) = \mathbb{P}(\mathcal{A}_g \subset [k] \mid A_0 = n),$$

where the first equality follows from Remark 3.5.2, the second one from (3.23) and the third one from Remark 3.5.5. \square

Another consequence of (3.23) together with Remark 3.5.5 is the following moment duality, which is interesting in its own right, not least because this was the route through which [GS18] discovered the ‘‘discrete ancestral selection graph’’ in the ‘‘quasi Wright-Fisher case’’, i.e. for $\mathbb{P}(W_1 = \dots = W_N) \rightarrow 1$.

Corollary 3.5.6. *Let (K_g) and (A_m) be as in Section 3.3.1, let $k, n \in [N]$ and assume that the number of wildtype individuals in generation 0 is k . Then the probability that a sample of n individuals taken in generation $g \geq 1$ consists of wildtype individuals only is*

$$\mathbb{E} \left[\binom{K_{g-1}}{\sum_{i=1}^n W_i} \mid K_0 = k \right] = \mathbb{E} \left[\binom{k}{\sum_{i=1}^n W_i} \mid A_0 = n \right],$$

where $G^{(1)}, G^{(2)}, \dots$ are independent and $\text{Geom}(1 - s_N)$ -distributed.

3.6 Coupling of the Cannings and Moran ancestral selection processes. Proof of Theorem 8a

In this section we provide a few lemmata preparing the proof of part a) of Theorem 8, and conclude with the proof of that part. In particular, in Lemma 3.6.9 we give a coupling of the Cannings ancestral selection process, for short CASP, $(A_m)_{m \geq 0}$ defined in Section 3.2.3 and the Moran ancestral selection process, for short MASP, $(B_r)_{r \geq 0}$ whose jump rates we recalled in Section 3.4.

Assume throughout that the Δ_N -valued random weights $\mathscr{W}^{(N)} = (W_1^{(N)}, \dots, W_N^{(N)})$ fulfill the assumptions (3.10) and (3.11) required in the first part of Theorem 8. Let $(s_N)_{N \geq 0}$ be a sequence in $(0, 1)$ obeying (3.2). Frequently, we will switch to the notation

$$b_N := -\frac{\ln s_N}{\ln N} \quad \text{or equivalently to} \quad s_N = N^{-b_N} \quad (3.25)$$

with (3.2) translating into

$$\frac{2}{3} + \eta \leq b_N \leq 1 - \eta.$$

For fixed N , and $j \in [N]$ let

$$G^{(j)} \text{ be independent and } \text{Geom}(1 - s_N)\text{-distributed;}$$

these will play the role of the random variables $G(j, g)$ defined in (3.18), see also (3.24). (Here and whenever there is no danger of confusion, we will suppress the superscripts N and g .)

Lemma 3.6.1 (Moran-like transition probabilities of the CASP).

Let $\varepsilon \in (0, \frac{1}{6})$. The transition probabilities of the CASP $(A_m)_{m \geq 0} = (A_m^{(N)})_{m \geq 0}$ obey, uniformly in $k \leq N^{1-b_N+\varepsilon}$,

$$\mathbb{P}(A_{m+1} = k | A_m = k) = 1 - ks_N - \binom{k}{2} \frac{\rho^2}{N} + O(k^4 N^{-2} + k^2 s_N^2) \quad (3.26)$$

$$\mathbb{P}(A_{m+1} = k + 1 | A_m = k) = ks_N + O(k^2 s_N^2 + k^4 N^{-2}) \quad (3.27)$$

$$\mathbb{P}(A_{m+1} = k - 1 | A_m = k) = \binom{k}{2} \frac{\rho^2}{N} + O(k^4 N^{-2} + k^2 s_N^2) \quad (3.28)$$

$$\mathbb{P}(|A_{m+1} - k| \geq 2 | A_m = k) = O(k^4 N^{-2} + k^2 s_N^2). \quad (3.29)$$

Remark 3.6.2. For $k = 2$ we have by (3.28) and (3.10)

$$\mathbb{P}(A_{m+1} = k - 1 | A_m = k) = \mathbb{E} \left[\sum_{i=1}^N \frac{W_i^2}{N} \right] = \frac{\rho^2}{N} + O(N^{-2}),$$

where the first term on the r.h.s. is the pair coalescence property of the neutral Cannings coalescent with the paintbox \mathscr{W} .

Proof of Lemma 3.6.1. Recall that each transition of the CASP consists of a branching and a coalescence step. To arrive at the transition probabilities (3.26) - (3.29) we first estimate the probabilities that k individuals give rise to a total of k , $k + 1$ or more than $k + 1$ branches and then analyse the probabilities that a single individual is chosen multiple times as a parent.

Since each individual has a $\text{Geom}(1 - s_N)$ -distributed number of branches, the probability that k individuals give rise to a total of k branches in the branching step

is

$$\mathbb{P} \left(\sum_{j=1}^k G^{(j)} = k \right) = (1 - s_N)^k = 1 - ks_N + O(k^2 s_N^2) \quad (3.30)$$

and the probability that the individuals give rise to $k + 1$ branches is

$$\mathbb{P} \left(\sum_{j=1}^k G^{(j)} = k + 1 \right) = k(1 - s_N)^k s_N = ks_N + O(k^2 s_N^2). \quad (3.31)$$

Adding the probabilities in (3.30) and (3.31) yields

$$\mathbb{P} \left(\sum_{j=1}^k G^{(j)} \geq k + 2 \right) = O(k^2 s_N^2).$$

Let us now calculate the probabilities of collisions in a coalescence step, that is the probability that an individual is chosen as a potential parent more than once. For two branches the pair coalescence probability c_N is given by

$$c_N = \mathbb{E} \left[\sum_{i=1}^N W_i^2 \right] = \frac{\rho^2}{N} + O(N^{-2}). \quad (3.32)$$

In the same manner we obtain the probability for a triple collision as

$$d_N = \mathbb{E} \left[\sum_{i=1}^N W_i^3 \right] = O(N^{-2}). \quad (3.33)$$

Using (3.32) and (3.33) we control the probability of the event E that there are two or more collisions, with k individuals before the coalescence step. There are two possibilities for this event to occur, either there is at least a triple collision or there are at least two pair collisions. This yields

$$\mathbb{P}(E) \leq \binom{k}{4} \frac{\rho^4}{N^2} + O \left(\binom{k}{3} N^{-2} \right) + O(k^4 N^{-3}) = O(k^4 N^{-2}). \quad (3.34)$$

In order to estimate the probability of having exactly one collision we use the second moment method for the random variable $X = \sum_{i=1}^k \sum_{j>i}^k X_{i,j}$, where $X_{i,j} = \mathbb{1}_{\{i \text{ and } j \text{ collide}\}}$. With (3.32) we get

$$\mathbb{E}[X] = \mathbb{E} \left[\sum_{i=1}^k \sum_{j>i}^k X_{i,j} \right] = \binom{k}{2} \frac{\rho^2}{N} + O(k^2 N^{-2}). \quad (3.35)$$

Furthermore, the second moment of X can be written again due to (3.32) and (3.33) as

$$\begin{aligned}\mathbb{E}[X^2] &= \mathbb{E}\left[\left(\sum_{i=1}^k \sum_{j>i}^k X_{i,j}\right)^2\right] \\ &= \binom{k}{2} \left(\frac{\rho^2}{N} + O(N^{-2})\right) + O(k^3 \mathbb{E}[X_{1,2}X_{2,3}]) + O(k^4 \mathbb{E}[X_{1,2}X_{3,4}]) \\ &= \binom{k}{2} \frac{\rho^2}{N} + O(k^3 N^{-2}) + O(k^4 N^{-2}) = \binom{k}{2} \frac{\rho^2}{N} + O(k^4 N^{-2})\end{aligned}$$

This together with (3.35) yields

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} = \frac{\left(\binom{k}{2} \frac{\rho^2}{N}\right)^2 + O(k^4 N^{-3})}{\binom{k}{2} \frac{\rho^2}{N} + O(k^4 N^{-2})} = \binom{k}{2} \frac{\rho^2}{N} (1 - O(k^2 N^{-1})),$$

where the first inequality follows by applying the Cauchy-Schwarz to X and $I_{\{X>0\}}$. Together with (3.34) we obtain for the random variable X which counts the number of collisions (for k individuals before the coalescence step)

$$\begin{aligned}\mathbb{P}(X = 0) &= 1 - \binom{k}{2} \frac{\rho^2}{N} + O(k^4 N^{-2}) \\ \mathbb{P}(X = 1) &= \binom{k}{2} \frac{\rho^2}{N} + O(k^4 N^{-2}) \\ \mathbb{P}(X \geq 2) &= \mathbb{P}(E) = O(k^4 N^{-2})\end{aligned}$$

Let $H := \sum_{j=1}^{A_g} G^{(j)}$. Then the above calculations allow us to obtain (3.26):

$$\begin{aligned}\mathbb{P}(A_{m+1} = k+1 | A_m = k) &= \mathbb{P}(A_{m+1} = k+1 | A_m = k, H = k) \mathbb{P}(H = k | A_m = k) \\ &\quad + \mathbb{P}(A_{m+1} = k+1 | A_m = k, H = k+1) \mathbb{P}(H = k+1 | A_m = k) \\ &\quad + \mathbb{P}(A_{m+1} = k+1 | A_m = k, H \geq k+2) \mathbb{P}(H \geq k+2 | A_m = k) \\ &= \left(1 - \binom{k+1}{2} \frac{\rho^2}{N} + O(k^4 N^{-2})\right) (ks_N + O(k^2 s_N^2)) + O(k^2 s_N^2) \\ &= ks_N + O(k^2 s_N^2 + s_N k^5 N^{-2})\end{aligned}$$

The remaining transition probabilities (3.27) - (3.29) are derived analogously. \square

The next lemma controls the speed of convergence to 1 of the probability of the event that the CASP comes down from N to (the still large state) $N^{1-b+\varepsilon}$ within a time interval of length $o(N^b)$. The quantities b, A_m, τ appearing in this lemma all depend on N ; we will suppress this dependence in the notation.

Lemma 3.6.3 (CASP coming down from huge to large). *Let $(A_m)_{m \in \mathbb{N}_0}$ be a CASP, $0 < \varepsilon < \frac{2}{3}$, $A_0 = N$ and denote by $\tau = \inf \{m \geq 0 : A_m \leq N^{1-b+\varepsilon}\}$ the first time the CASP crosses the level $N^{1-b+\varepsilon}$. Then there exists a $\delta > 0$, such that for any constant $c > 0$*

$$\mathbb{P}\left(\frac{\tau}{N^b} > c\right) = O(\exp(-N^\delta)). \quad (3.36)$$

Proof. The branching step of the CASP dynamics only depends on s_N and neither on the distribution nor the realization of \mathscr{W} , see Remark 3.5.5 i). On the other hand, the coalescence step only depends on \mathscr{W} . In Lemma 3.6.4 below we will show that the size of this coalescence step is stochastically dominated from below by the corresponding step in a Wright-Fisher coalescent. Thus, among all the CASP's with selective strength s_N , the CASP of the Wright-Fisher model with selection is the slowest to come down from N to $N^{1-b+\varepsilon}$; therefore we use the stopping time corresponding to the Wright-Fisher model as a stochastic upper bound for τ . Consequently, we assume for the rest of the proof that $\mathscr{W} = (\frac{1}{N}, \dots, \frac{1}{N})$.

To show (3.36) we estimate $\mathbb{E}[A_{m+1}|A_m = k]$ for $1 \leq k \leq N$.

A_{m+1} denotes the number of potential parents of A_m individuals, that is

$$A_{m+1} = \sum_{i=1}^N \mathbb{1}_{\{\text{Individual } i \text{ is a potential parent of some of the } A_m \text{ individuals}\}}.$$

Let $H = \sum_{j=1}^{A_m} G^{(j)}$, with $G^{(j)} \sim \text{Geom}(1 - s_N)$ and independent for $j \in [N]$. Then

$$\mathbb{P}(\text{Individual } i \text{ is chosen as a potential parent} | A_m) = 1 - \mathbb{E} \left[\left(1 - \frac{1}{N}\right)^H \mid A_m \right]$$

for $i \in [N]$. Hence, for $k \geq 1$ and $x = \frac{1}{N(1-s_N)}$

$$\begin{aligned} \mathbb{E}[A_{m+1}|A_m = k] &= N \mathbb{E} \left[\left(1 - \left(1 - \frac{1}{N}\right)^H\right) \mid A_m = k \right] \\ &= N \left(1 - \left(1 - \frac{1}{N(1-s_N + \frac{s_N}{N})}\right)^k\right) \\ &\leq -N \left(k \ln(1-x) + \frac{(k \ln(1-x))^2}{2} + \frac{(k \ln(1-x))^3}{6} \right) \\ &\leq N \left(kx - \frac{(kx)^2(1-2s_N)}{3} \right) \\ &= \frac{k}{1-s_N} - \frac{k^2(1-2s_N)}{3N}, \end{aligned} \tag{3.37}$$

where for (3.37) we use the probability generating function of the negative binomial distribution and for the next line we use an estimate for the remainder of the corresponding Taylor expansion. Let $0 < \varepsilon' < \varepsilon$. From (3.38) follows

$$\begin{aligned} \mathbb{E}[A_{m+1}|A_m] &\leq \max \left\{ \frac{A_m}{1-s_N} - \frac{A_m N^{1-b+\varepsilon'}(1-2s_N)}{3N}, N^{1-b+\varepsilon'} \right\} \\ &= \max\{q_N A_m, N^{1-b+\varepsilon'}\}, \end{aligned}$$

with $q_N = \frac{1}{1-s_N} - \frac{N^{1-b+\varepsilon'}(1-2s_N)}{3N}$. This yields

$$\mathbb{E}[A_m|A_0 = N] \leq \max\{q_N^m N, N^{1-b+\varepsilon'}\}.$$

For any $m \geq c_1 N^{b-\varepsilon'} \ln N$, for some appropriate constant $c_1 > 0$ we have thus the estimate $\mathbb{E}[A_m | A_0 = N] \leq N^{1-b+\varepsilon}$. By Markov's inequality we obtain

$$\mathbb{P}(A_{c_1 N^{b-\varepsilon'} \ln N} > N^{1-b+\varepsilon}) \leq N^{\varepsilon'-\varepsilon} \rightarrow 0$$

as $N \rightarrow \infty$. If $(A_m)_{m \geq 0}$ did not reach $N^{1-b+\varepsilon}$ after $c_1 N^{b-\varepsilon'} \ln N$ steps we can start the process in N again and wait another $c_1 N^{b-\varepsilon'} \ln N$ steps and check whether the process did reach the level $N^{1-b+\varepsilon}$. By using this argument N^{δ_1} times this yields, for any $0 < \delta_1 < \varepsilon'$, the following upper bound for the probability to stay above $N^{1-b+\varepsilon}$ for the generations $m \leq c_1 N^{b-\varepsilon'+\delta_1}$:

$$\begin{aligned} \mathbb{P}(A_m > N^{1-b+\varepsilon} \text{ for } m \in \{0, \dots, c_1 N^{b-\varepsilon'+\delta_1} \ln N\}) &\leq \mathbb{P}(A_{c_1 N^{b-\varepsilon'} \ln N} > N^{1-b+\varepsilon})^{N^{\delta_1}} \\ &\leq (N^{\varepsilon'-\varepsilon})^{N^{\delta_1}}. \end{aligned}$$

Since $\varepsilon > \varepsilon'$ and $N^{b-\varepsilon'+\delta_1} < N^b$, we have $(N^{\varepsilon'-\varepsilon})^{N^{\delta_1}} = O(\exp(-N^\delta))$ for some appropriate $\delta > 0$ from which the assertion follows. \square

We now append the lemma on the extremality of the Wright-Fisher coalescent which we used in the previous proof, and put this lemma right away in the context of the "coupon collector".

Lemma 3.6.4. *For natural numbers N and k let Z_1, Z_2, \dots , be i.i.d. $[N]$ -valued random variables with $w_i := \mathbb{P}(Z_1 = i)$, $i \in [N]$. Then for each $k \in \mathbb{N}$ the random variable $C_k := |\{Z_1, \dots, Z_k\}|$ is stochastically largest for $w_1 = \dots = w_N = \frac{1}{N}$.*

Proof. For $\ell \in [N]$ we consider $T_\ell := \min\{k : C_k \geq \ell\}$. Then we have the obvious identity

$$\{C_k \geq \ell\} = \{T_\ell \leq k\}.$$

[ABSS16, Theorem 2] states that $\mathbb{P}(T_\ell \leq k)$ is largest for $w_1 = \dots = w_N = \frac{1}{N}$. \square

From Lemma 3.6.3 we obtain the following corollary

Corollary 3.6.5. *Let $(A_m)_{m \geq 0}$ be a CASP. Then for any $m_0 \geq 0$ there exists a $C > 0$ such that for all $N \geq 1$ and all $j \geq N^{1-b+\varepsilon}$*

$$\begin{aligned} \mathbb{P}(A_{m_0+N^b} > j | A_0 = N) &\leq C N^{1-b+\varepsilon} / j, \\ \mathbb{E}[A_{m_0+N^b} | A_0 = N] &= O(\ln(N) N^{1-b+\varepsilon}). \end{aligned}$$

Proof. For simplicity assume that $m_0 = 0$, but the same proof works for any $m_0 \in \mathbb{N}$ as $\mathbb{P}(A_{m_0+N^b} > j | A_0 = N) \leq \mathbb{P}(A_{m_0+N^b} > j | A_{m_0} = N)$. Due to Lemma 3.6.3 for the stopping time $\tau = \inf\{m \geq 0 : A_m \leq N^{1-b+\varepsilon}\}$ it holds $\mathbb{P}(\tau > N^b) = O(\exp(-N^{-\delta}))$, with δ as in Lemma 3.6.3. By Lemma 3.6.1 we can compare the jump probabilities and obtain that there exists some $x_0 \leq N^{1-b+\varepsilon/2}$ such that above x_0 the upward drift is smaller than the downward drift. This yields that the process stopped in x_0 is a supermartingale. Consequently since $x_0 < N^{1-b+\varepsilon}$, we have for any $m' \in \mathbb{N}$ by the strong Markov property

$$\mathbb{E}[A_{\tau+m'}] \leq N^{1-b+\varepsilon}.$$

Hence by Markov's inequality we obtain

$$\begin{aligned} \mathbb{P}(A_{N^b} > j | A_0 = N) &\leq \mathbb{P}(A_{N^b} > j | A_0 = N, \tau \leq N^b) + \mathbb{P}(\tau > N^b) \\ &\leq \frac{\mathbb{E}[A_{N^b} | A_0 = N, \tau \leq N^b]}{j} + O(\exp(-N^\delta)) \\ &\leq \frac{\mathbb{E}[A_\tau]}{j} + O(\exp(-N^\delta)), \end{aligned}$$

which shows the first part. For the second part observe that

$$\begin{aligned} \mathbb{E}[A_{N^b} | A_0 = N] &= \sum_{j=1}^N \mathbb{P}(A_{N^b} > j | A_0 = N) \\ &= \sum_{j=1}^{N^{1-b+\varepsilon}} \mathbb{P}(A_{N^b} > j | A_0 = N) + \sum_{j=N^{1+b-\varepsilon}}^N \mathbb{P}(A_{N^b} > j | A_0 = N) \\ &\leq N^{1-b+\varepsilon} + \sum_{j=N^{1+b-\varepsilon}}^N \frac{N^{1-b+\varepsilon}}{j} = O(N^{1-b+\varepsilon} \ln N). \end{aligned}$$

□

The following three lemmata provide some properties about the Moran process and the coupling of a Moran process to a Moran process in stationarity. For the remainder of this section we will fix three constants

$$\delta_1 \in (0, 1), \quad 0 < \delta_3 < \delta_2/2. \quad (3.39)$$

The role of δ_1 will be to specify a region $\left[\frac{2s_N}{2s_N+\rho^2} N(1-\delta_1), \frac{2s_N}{2s_N+\rho^2} N(1+\delta_1) \right]$ around MASP's center of attraction. The constant δ_2 will appear in factors N^{δ_2} that stretch some time intervals, and the constant δ_3 will be an exponent in small probabilities $O(\exp(-N^{\delta_3}))$.

Lemma 3.6.6 (MASP's hitting time of the central region). *Let $(B_r)_{r \geq 0}$ be a MASP started in some state $n \in [N]$ and let $T = \inf\{r \geq 0 : B_r \in [\frac{2s_N}{2s_N+\rho^2} N(1-\delta_1), \frac{2s_N}{2s_N+\rho^2} N(1+\delta_1)]\}$. Then,*

$$\mathbb{P}(T \leq N^{b+\delta_2} | B_0 = n) = 1 - O(\exp(-N^{\delta_3})).$$

Proof. We proceed in a similar manner as [PP13] and separate the proof into two cases

$$\text{i) } B_0 > \frac{2s_N}{2s_N+\rho^2} N(1+\delta_1)$$

$$\text{ii) } B_0 < \frac{2s_N}{2s_N+\rho^2} N(1-\delta_1).$$

For case i) the proof relies on a stochastic domination of the MASP by a birth-death process, while for case ii) we construct a pure birth process that is stochastically dominated by the MASP. We start by proving case i).

Assume the most extremal starting point $B_0 = N$. We couple the process $(B_r)_{r \geq 0}$ with a birth-death process $(\bar{B}_r)_{r \geq 0}$ which stochastically dominates $(B_r)_{r \geq 0}$ until $(B_r)_{r \geq 0}$ crosses the level $\frac{2s_N}{2s_N+\rho^2} N(1+\delta_1)$. $(\bar{B}_r)_{r \geq 0}$ is defined as the Markov process with state space \mathbb{N}_0 and the following transition rates

- $k \rightarrow k + 1$ with rate $ks_N =: \bar{\beta}k$
- $k \rightarrow k - 1$ with rate $k \frac{s_N \rho^2}{2s_N + \rho^2} (1 + \delta_1) =: \bar{\alpha}k$.

Note that $\bar{\beta}k \geq s_N k(N - k)/N$ and $\bar{\alpha}k \leq \binom{k}{2} \frac{\rho^2}{N}$ for any $k \geq \frac{2s_N}{2s_N + \rho^2} N(1 + \delta_1)$. Hence, we can couple $(\bar{B}_r)_{r \geq 0}$ and $(B_r)_{r \geq 0}$ such that $B_r \leq \bar{B}_r$ a.s. as long as $B_r \geq \frac{2s_N}{2s_N + \rho^2} N(1 + \delta_1)$. In particular, we have

$$\mathbb{P}(T \geq r | B_0 = k) \leq \mathbb{P}(\bar{\tau}_0 \geq r | \bar{B}_0 = k) \quad (3.40)$$

when we set $\bar{\tau}_0 := \inf \{r \geq 0 : \bar{B}_r = 0\}$ and $k \geq \frac{2s_N}{2s_N + \rho^2} N(1 + \delta_1)$. For the birth-death process \bar{B}_r we can estimate $\bar{\tau}_0$, by a classical *first step* analysis

$$\begin{aligned} \mathbb{P}(\bar{\tau}_0 \geq r | \bar{B}_0 = 1) &= (1 - (\bar{\alpha} + \bar{\beta})dr) \mathbb{P}(\bar{\tau}_0 \geq r - dr | \bar{B}_0 = 1) \\ &\quad + \bar{\beta}dr(1 - (1 - \mathbb{P}(\bar{\tau}_0 \geq r - dr | \bar{B}_0 = 1))^2) \end{aligned}$$

Setting $f(r) = \mathbb{P}(\bar{\tau}_0 \geq r | \bar{B}_0 = 1)$ we obtain

$$f'(r) = (\bar{\beta} - \bar{\alpha})f(r) - \bar{\beta}f(r)^2$$

with $f(0) = 1$ which is solved by

$$f(r) = \frac{\bar{\alpha} - \bar{\beta}}{\bar{\alpha}e^{r(\bar{\alpha} - \bar{\beta})} - \bar{\beta}}.$$

Observe that $\bar{\alpha} - \bar{\beta} = \frac{s_N \rho^2}{\rho^2 + 2s_N} (1 + \delta_1) - s_N = \delta_1 s_N (1 + o(1))$, hence

$$\begin{aligned} f(N^{b+\delta_2}) &= \frac{\delta_1 s_N (1 + o(1))}{\frac{\rho^2 s_N}{2s_N + \rho^2} (1 + \delta_1) \exp(N^{b+\delta_2} \delta_1 s_N (1 + o(1))) - s_N} \\ &= \frac{\delta_1 (1 + o(1))}{\frac{\rho^2}{2s_N + \rho^2} (1 + \delta_1) \exp(N^{\delta_2} \delta_1 (1 + o(1))) - 1}. \end{aligned} \quad (3.41)$$

From (3.40) and (3.41) we finally estimate

$$\begin{aligned} \mathbb{P}(T < N^{b+\delta_2} | B_0 = N) &\geq \mathbb{P}(\bar{\tau}_0 < N^{b+\delta_2} | \bar{B}_0 = N) \geq 1 - Nf(N^{b+\delta_2}) \\ &= 1 - O(N \exp(-N^{\delta_2})) = 1 - O(\exp(-N^{\delta_3})) \end{aligned}$$

for any $\delta_3 < \delta_2$. This proves part i).

Now it remains to prove the case ii). Again assume the most extremal starting point $B_0 = 1$. Let $(\underline{B}_r)_{r \geq 0}$ be a birth-death process which jumps

- from k to $k + 1$ at rate $ks_N(1 - \frac{2s_N}{2s_N + \rho^2}(1 - \delta_1)) =: \underline{\beta}k$
- from k to $k - 1$ at rate $k \frac{s_N \rho^2}{2s_N + \rho^2} (1 - \delta_1) =: \underline{\alpha}k$.

Observe that $\underline{\beta}k \leq s_N k(N - k)/N$ and $\underline{\alpha}k \geq \binom{k}{2} \frac{\rho^2}{N}$ as long as $k \leq \frac{2s_N}{\rho^2 + 2s_N} N(1 - \delta_1)$. Hence, we can couple $(\underline{B}_r)_{r \geq 0}$ and $(B_r)_{r \geq 0}$ such that $B_r \geq \underline{B}_r$ as long as $B_r \leq \frac{2s_N}{\rho^2 + 2s_N} N(1 - \delta_1)$.

The extinction probability ξ_0 of $(\underline{B}_r)_{r \geq 0}$ is the smallest solution of

$$\xi = \frac{\underline{\beta}}{\underline{\beta} + \underline{\alpha}} \xi^2 + \frac{\underline{\alpha}}{\underline{\beta} + \underline{\alpha}},$$

that is $\xi_0 = \frac{\underline{\alpha}}{\underline{\beta}} < 1$. Let $(\underline{B}_r^I)_{r \geq 0}$ be the pure birth process consisting of the immortal lines of $(\underline{B}_r)_{r \geq 0}$, i.e. each line branches at rate $(1 - \xi_0)\underline{\beta}$.

Let $\tau = \inf\{r \geq 0 : B_r \geq \frac{2}{\rho^2} s_N N(1 - \delta_1)\}$ be the time when $(B_r)_{r \geq 0}$ reaches the level $\frac{2}{\rho^2} s_N N(1 - \delta_1)$ and define $\underline{\tau}^I$ and $\underline{\tau}$ in the same way for the processes $(\underline{B}_r^I)_{t \geq 0}$ and $(\underline{B}_r)_{r \geq 0}$ respectively in place of $(B_r)_{r \geq 0}$, then $\underline{\tau}^I \geq \underline{\tau} \geq \tau$ a.s. In order to prove ii) it remains to show $\mathbb{P}(\underline{\tau}^I \geq N^{b+\delta_2}) = O(\exp(-N^{\delta_3}))$ for $\delta_3 > 0$. We have

$$\begin{aligned} \mathbb{E}[\underline{\tau}^I] &= \mathbb{E} \left[\sum_{i=1}^{\lfloor \frac{2s_N}{2s_N + \rho^2} N(1 - \delta_1) \rfloor} \frac{1}{i\underline{\beta}(1 - \xi_0)} \right] = \frac{1}{\underline{\beta}(1 - \xi_0)} \left(\ln \left(\frac{2s_N(1 - \delta_1)}{2s_N + \rho^2} N \right) + O(1) \right) \\ &= \frac{1}{\delta_1 s_N} \left(\ln \left(\frac{2s_N(1 - \delta_1)}{2s_N + \rho^2} N \right) + O(1) \right) \\ &= \frac{1}{\delta_1} N^b \ln \left(\frac{2(1 - \delta_1)}{\rho^2} N^{1-b} \right) (1 + O(s_N)) \\ &= \frac{1-b}{\delta_1} N^b \ln(N) (1 + O((\ln N)^{-1})). \end{aligned}$$

We can estimate $\mathbb{P}(\underline{\tau}^I > N^{b+\delta_2} | \underline{B}_0^I = 1) \leq \mathbb{P}_1(\underline{\tau}^I > N^{b+\delta_2/2} | \underline{B}_0^I = 1)^{N^{\delta_2/2}}$ for $\delta_2 > 0$ by separating the time interval of length $N^{b+\delta_2}$ into $N^{\delta_2/2}$ time intervals of length $N^{b+\delta_2/2}$ and realizing that if $(\underline{B}_r^I)_{t \geq 0}$ did not reach the level $\frac{2}{\rho^2} s_N N(1 - \delta_1)$ in a time interval of length $N^{b+\delta_2/2}$ then in the worst case $(\underline{B}_r^I)_{r \geq 0}$ is 1 at the start of each time interval.

By Markov's inequality we then arrive at

$$\begin{aligned} \mathbb{P}(\underline{\tau}^I \geq N^{b+\delta_2}) &\leq \mathbb{P}(\underline{\tau}^I > N^{b+\frac{\delta_2}{2}})^{N^{\frac{\delta_2}{2}}} \leq \left(\frac{1}{\delta_1} N^{-\frac{\delta_2}{2}} \ln N \right)^{N^{\frac{\delta_2}{2}}} \\ &= \exp \ln \left(\left(\frac{1}{\delta_1} N^{-\frac{\delta_2}{2}} \ln N \right)^{N^{\frac{\delta_2}{2}}} \right) \\ &\leq \exp \left(-\frac{\delta_2}{2} N^{\frac{\delta_2}{2}} \right) = O(\exp(-N^{\delta_3})) \end{aligned} \quad (3.42)$$

for $\delta_3 < \delta_2/2$. From (3.42) we can directly conclude $\mathbb{P}(\tau \geq N^{b+\delta_2}) = O(\exp(-N^{\delta_3}))$, which together with part i) finishes the proof. \square

Lemma 3.6.7 (MASP's leaving time of the central region). *Let $(B_r)_{r \geq 0}$ be a MASP started in $x \in [\frac{2s_N}{2s_N + \rho^2} N(1 - \delta_1), \frac{2s_N}{2s_N + \rho^2} N(1 + \delta_1)]$ and assume in addition to (3.39) that $0 < \delta_1 < \frac{1}{2}$ and $0 < \delta_2 < \frac{\eta}{3}$. Let $S = \inf\{r \geq 0 : B_r \notin [\frac{2s_N}{2s_N + \rho^2} N(1 - 2\delta_1), \frac{2s_N}{2s_N + \rho^2} N(1 + 2\delta_1)]\}$. Then*

$$\mathbb{P}(S > N^{b+\delta_2}) \geq 1 - o(\exp(-N^{1-b-3\delta_2})). \quad (3.43)$$

Proof. Assume we have $B_0 \in [\frac{2s_N}{2s_N+\rho^2}N(1-\delta_1), \frac{2s_N}{2s_N+\rho^2}N(1+\delta_1)]$. To prove (3.43) we couple $(B_r)_{r \geq 0}$ with a symmetric (discrete time) random walk $(S_n)_{n \geq 0}$, and thus ignore the drift to $\frac{2s_N}{2s_N+\rho^2}N$. An application of Theorem 5.1 iii) of [Jan18] yields that $(B_r)_{r \geq 0}$ makes at most $N^{1-b+2\delta_2}$ many jumps in a time interval of length $N^{b+\delta_2}$ with probability $1 - O(\exp(-N^{1-b+2\delta_2}))$, see also the estimate (3.47) in Lemma 3.6.9 below, where we analyze the jumps and jump times of the MASP in more detail. Hence,

$$\begin{aligned} & \mathbb{P}_x \left(B_r \notin \left[\frac{2s_N}{2s_N+\rho^2}N(1-2\delta_1), \frac{2s_N}{2s_N+\rho^2}N(1+2\delta_1) \right] \text{ for some } r \leq N^{b+\delta_2} \right) \\ & \leq \mathbb{P}_0 \left(S_n \notin \left[-\delta_1 \frac{2s_N}{2s_N+\rho^2}N, \delta_1 \frac{2s_N}{2s_N+\rho^2}N \right] \text{ for some } n \leq N^{1-b+2\delta_2} \right) \\ & = 2\mathbb{P}_0 \left(\max_{1 \leq n \leq N^{1-b+2\delta_2}} S_n \notin \left[0, \delta_1 \frac{2s_N}{2s_N+\rho^2}N \right] \right) \\ & = 4\mathbb{P}_0 \left(S_{N^{1-b+2\delta_2}} > \delta_1 \frac{2s_N}{2s_N+\rho^2}N \right) \tag{3.44} \end{aligned}$$

$$\leq 4 \exp \left(-cN^{1-b-2\delta_2} \right) = o(\exp(-N^{1-b-3\delta_2})) \tag{3.45}$$

for some appropriate $c > 0$ independent of N . To obtain equation (3.44) and inequality (3.45) we used the reflection principle and Hoeffding's inequality. This finishes the proof. \square

Lemma 3.6.8 (MASP close to stationarity). *Let $(B_r)_{r \geq 0}$ be a MASP started in k individuals, with $1 \leq k \leq N$, then*

$$d_{\text{TV}}(\mathcal{L}(B_{N^{b+\delta_2}}), \mathcal{L}(B_{\text{eq}})) = O(\exp(-N^{\delta_3}))$$

with $B_{\text{eq}} = B_{\text{eq}}^{(N)}$ as in (3.16), i.e. distributed as a Binomial($N, \frac{2s_N}{2s_N+\rho^2}$)-random variable conditioned to be strictly positive, and the constant in the Landau O is uniform in k .

Proof. We follow a similar strategy as the one used in the proof of Lemma 2.10 in [PP13]. Let $(B_r^{\text{eq}})_{r \geq 0}$ be a MASP started in the stationary distribution. Assume that in the graphical representation at time 0 either the lines of B_0 are contained in B_0^{eq} or vice versa. Then $B_r \leq B_r^{\text{eq}}$, for all $r \geq 0$, or vice versa $B_r^{\text{eq}} \leq B_r$. Then $\mathbb{P}(B_{N^{b+\delta_2}} = k) = \mathbb{P}(B^{\text{eq}} = k)(1 - O(e^{-N^{\delta_3}}))$ follows, once we show that at time $N^{b+\delta_2}$ both processes are equal with probability $(1 - O(e^{-N^{\delta_3}}))$.

The tuple $(B_r^{\text{eq}}, B_r)_{r \geq 0}$, and the tuple $(B_r, B_r^{\text{eq}})_{r \geq 0}$ resp., have the following transition rates: jumps from (k, ℓ) for $1 \leq k \leq \ell \leq N$ to

- $(k+1, \ell+1)$ occur at rate $s_N k(1 - \frac{\ell}{N})$
- $(k, \ell+1)$ occur at rate $s_N(\ell - k)(1 - \frac{\ell}{N})$
- $(k+1, \ell)$ occur at rate $ks_N \frac{\ell-k}{N}$
- $(k, \ell-1)$ occur at rate $\frac{\rho^2}{N} \left(\binom{\ell-k}{2} + (\ell-k)k \right)$
- $(k-1, \ell-1)$ occur at rate $\frac{\rho^2}{N} \binom{k}{2}$.

To proceed further we consider the two cases

- i) $B_0 > B_0^{\text{eq}}$
- ii) $B_0 < B_0^{\text{eq}}$

separately.

We begin with Case i). Consider the process $(Z_r)_{r \geq 0}$ defined as $Z_r := B_r - B_r^{eq}$ and condition on the two events that the process B_0^{eq} is started in a state in $[\frac{2s_N}{2s_N + \rho^2}N(1 - \delta_1), \frac{2s_N}{2s_N + \rho^2}N(1 + \delta_1)]$ and stays in $[\frac{2s_N}{2s_N + \rho^2}N(1 - 2\delta_1), \frac{2s_N}{2s_N + \rho^2}N(1 + 2\delta_1)]$ for some $0 < \delta_1 < \frac{1}{2}$. The probability of each event can be estimated by $1 - O(\exp(-N^{\delta_2}))$, the former event by Hoeffding's inequality and the latter with Lemma 3.6.7. The process $(Z_r)_{r \geq 0}$ jumps from z to $z+1$ at most at rate $s_N z$ and under the above condition $(Z_r)_{r \geq 0}$ jumps from z to $z-1$ at least at rate $\rho^2 \frac{2s_N}{2s_N + \rho^2} (1 - 2\delta_1) z$: If $(Z_r, B_r, B_r^{eq}) = (z, \ell, k)$ jumps to $(z-1, \ell-1, k)$ occur at rate $\frac{\rho^2}{N} \binom{z}{2} + zk$ and jumps to $(z-1, \ell, k+1)$ at rate $ks_N \frac{\ell-k}{N}$. Therefore, the process $(Z_r)_{r \geq 0}$ jumps from $z \rightarrow z-1$ at rate $r_{z, z-1} = \frac{\rho^2}{N} \binom{z}{2} + zk + ks_N \frac{z}{N}$. Due to the condition and the assumption that $\ell \geq k \geq \frac{2s_N}{2s_N + \rho^2}N(1 - 2\delta_1)$ we can bound

$$\begin{aligned} r_{z, z-1} &= \frac{\rho^2}{N} \left(\binom{z}{2} + zk \right) + ks_N \frac{z}{N} \geq \frac{\rho^2}{2N} z(k + \ell - 1) + z \frac{2s_N^2}{2s_N + \rho^2} (1 - 2\delta_1) \\ &\geq z \frac{\rho^2}{2N} 2 \frac{2s_N}{2s_N + \rho^2} N(1 - 2\delta_1) = z \rho^2 \frac{2s_N}{2s_N + \rho^2} (1 - 2\delta_1). \end{aligned}$$

Hence, we can couple $(Z_r)_{r \geq 0}$ to a birth-death process $(Z'_r)_{r \geq 0}$ with individual birth rate $s_N =: \beta'$ and individual death rate $\rho^2 \frac{2s_N}{2s_N + \rho^2} (1 - 2\delta_1) =: \alpha'$, such that $Z_r \leq Z'_r$ a.s. Let $\xi := \inf\{r \geq 0 : Z_r = 0\}$ and $\xi' := \inf\{r \geq 0 : Z'_r = 0\}$. Obviously it holds $\mathbb{P}(\xi \geq r) \leq \mathbb{P}(\xi' \geq r)$ for all $r \geq 0$. As in the proof of Lemma 3.6.6 we estimate

$$\begin{aligned} \mathbb{P}(\xi' \geq N^{b+\delta_2} | Z'_0 = 1) &= \frac{\left(\frac{2\rho^2(1-2\delta_1)}{2s_N + \rho^2} - 1 \right) s_N}{\left(\frac{2\rho^2(1-2\delta_1)}{2s_N + \rho^2} - 1 \right) s_N \exp\left(\left(\frac{2\rho^2(1-2\delta_1)}{2s_N + \rho^2} - 1 \right) N^{\delta_2} \right) - s_N} \\ &= O(\exp(-c_N N^{\delta_2})) \end{aligned}$$

with $c_N = \left(\frac{2\rho^2(1-2\delta_1)}{2s_N + \rho^2} - 1 \right) \rightarrow 2(1 - 2\delta_1) - 1 > 0$. Since $Z_0 \leq N$ the probability that all lines go extinct before time $N^{b+\delta_2}$ can be estimated by

$$\mathbb{P}(Z_{N^{b+\delta_2}} = 0) \geq \left(1 - \exp(-c_N N^{\delta_2}) \right)^N = 1 - O\left(\exp(-N^{\delta_3}) \right),$$

which proves Lemma 3.6.8 in Case i).

In Case ii) we first wait until $(B_r)_{r \geq 0}$ reaches the level $\frac{2}{\rho^2} s_N N(1 - \delta_1)$ within a time interval of length $O(N^{b+\delta_2})$ with probability $1 - O(\exp(-N^{\delta_3}))$ due to Lemma 3.6.6 and we assume that B_0^{eq} is started in at least $\frac{2}{\rho^2} s_N N(1 - \delta_1)$, which happens with probability $1 - O(\exp(-\delta_1^2 N))$ due to Hoeffding's inequality. Then due to Lemma 3.6.7 both processes remain bounded from below by $\frac{2}{\rho^2} s_N N(1 - 2\delta_1)$. When $(B_r)_{r \geq 0}$ has reached at least the level $\frac{2}{\rho^2} s_N N(1 - \delta_1)$ consider $Z_r = B_r - B_r^{eq}$. Then the same arguments as in Case i) show the claim. \square

As mentioned in the sketch of proof of Theorem 8 in Section 3.3 we aim to couple the CASP with the MASP. We have seen in the calculations before that in the regime where the number of potential ancestors is at most of order $N^{1-b+\varepsilon}$ for ε sufficiently small the transition probabilities of these two processes are essentially the same for a time interval of length of order $O(N^{b+\varepsilon})$. In particular in a time interval of length $O(N^{b+\varepsilon})$ we can exclude jumps of size 2 or bigger in the CASP with probability

$O(N^{-\delta})$.

Lemma 3.6.9 (Coupling of MASP and CASP). *Let $0 < \varepsilon < \frac{\eta}{2}$, and $0 < \delta = 3\eta - 6\varepsilon$. There exists a coupling of the MASP $(B_r)_{r \geq 0}$ and the CASP $(A_m)_{m \geq 0}$ such that for all common initial values k_0 with $1 \leq k_0 \leq N^{1-b+\varepsilon}$*

$$\mathbb{P}\left(|A_m - B_m| \leq 1, \forall m \in \{0, \dots, N^{b+\varepsilon}\}\right) = 1 - O(N^{-\delta}). \quad (3.46)$$

with the constant in the Landau O uniform in k_0 .

Proof. Let $A_0 = B_0 = k_0 \leq N^{1-b+\varepsilon}$. We will show that the CASP and the MASP can be coupled such that the jump times of the CASP and the MASP occur consecutively with probability $1 - O(N^{-\delta})$. Since the transition probabilities of the CASP and the MASP are essentially the same we can also couple the jump directions with high probability. To show that the jump times occur consecutively we first show the following claim.

Claim 1: With probability $1 - O(N^{-\delta})$ the MASP makes in each of the time intervals $[\ell - 1, \ell]$ at most one jump.

By Lemma 3.6.6 and 3.6.7 the MASP stays below $2N^{1-b+\varepsilon}$ with probability $1 - O(\exp(-N^{\delta_3}))$.

Denote by $r_{k,k+1}$ and $r_{k,k-1}$ the jump rates for the MASP from k to $k+1$ and from k to $k-1$ respectively with $\gamma = \rho^2$. Then

- $r_{k,k+1} = k s_N + O\left(\frac{k}{N}\right)$
- $r_{k,k-1} = \binom{k}{2} \frac{\rho^2}{N}$.

Define $r_k = r_{k,k+1} + r_{k,k-1}$ the total jump rate and

$$r_\star = r_{N^{1-b+\varepsilon}} = \max_{1 \leq k \leq N^{1-b+\varepsilon}} r_k = N^{1-2b+2\varepsilon} (1 + o(1))$$

the maximal jump rate. We aim for the coupling to hold for an interval of length $N^{b+\varepsilon}$. The jump times of $(B_r)_{r \geq 0}$ are exponentially distributed with a parameter bounded from above by r_\star . To estimate the number of jumps falling into an interval of length $N^{b+\varepsilon}$ we use Theorem 5.1 iii) in [Jan18]. Let $(X_i)_{i \geq 1}$ be a family of independent $\text{Exp}(r_\star)$ distributed random variables. For $c = 1 - b + 4\varepsilon$ Theorem 5.1 iii) yields

$$\mathbb{P}\left(\sum_{i=1}^{N^c} X_i \leq N^{b+\varepsilon}\right) = O(\exp(-N^{1-b+4\varepsilon})), \quad (3.47)$$

that is the number of jumps is bounded by $N^{1-b+4\varepsilon}$ with probability $1 - O(\exp(-N^{1-b}))$. For $E = \{(B_r)_{0 \leq r \leq N^b} \text{ has at most one jump in the intervals } [j, j+1] \text{ for each } 0 \leq j \leq N^b - 1\}$ we have

$$\begin{aligned} \mathbb{P}(E) &\geq \left(1 - O(\exp(-N^{1-b}))\right) \prod_{i=1}^{N^c} \mathbb{P}(X_i > 1) \geq \left(1 - O(\exp(-N^{1-b}))\right) e^{-r_\star N^c} \\ &= 1 - O(N^{-\delta}) \end{aligned}$$

which yields Claim 1.

Let $T_i^A = \inf\{m \geq T_{i-1}^A : A_m \neq A_{T_{i-1}^A}\}$ be the i -th jump of the CASP with the convention $T_{-1}^A = 0$. In the same manner let $T_i^B = \inf\{r \geq T_{i-1}^B : B_r \neq B_{T_{i-1}^B}\}$ be the

i -th jump of the MASP again with the convention that $T_{-1}^B = 0$. We have

$$\mathbb{P}(B_{T_i^B} = k+1 | B_{T_{i-1}^B} = k) = \frac{r_{k,k+1}}{r_{k,k+1} + r_{k,k-1}} = \mathbb{P}(A_{T_i^A} = k+1 | A_{T_{i-1}^A} = k) + e_{k,N}$$

and

$$\mathbb{P}(B_{T_i^B} = k-1 | B_{T_{i-1}^B} = k) = \frac{r_{k,k-1}}{r_{k,k+1} + r_{k,k-1}} = \mathbb{P}(A_{T_i^A} = k-1 | A_{T_{i-1}^A} = k) + f_{k,N},$$

where $e_{k,N}, f_{k,N} \in O(\max\{k^2 s_N^2, k^4 N^{-2}, N^{-1}\})$, the latter being the error terms from (3.27) and (3.28). Note that $e_{k,N}, f_{k,N} \geq 0$ because the CASP can make jumps of size 2 or larger. Set $d_{k,N} = e_{k,N} + f_{k,N}$.

We show that we can couple the times T_i^A and T_i^B , such that $T_{i+1}^B < T_i^A$ for $i = 1, \dots, N^{1-b+4\varepsilon}$ with probability $1 - O(N^{-\delta})$. From that follows the Assertion (3.46) of the Lemma by coupling the jump directions.

We couple the jump times T_i^A and T_i^B such that for all $i \in \{1, \dots, \lfloor N^{1-b+3\varepsilon} \rfloor\}$

$$\mathbb{P}(T_{i+1}^B < T_i^A) = O(N^{1-2b+2\varepsilon}) \quad (3.48)$$

from which follows the assertion. We explicitly construct the coupling for $i = 1$, and the same holds for any $i \in \{1, \dots, \lfloor N^{1-b+4\varepsilon} \rfloor\}$. To show (3.48) observe that, if $A_0 = k = B_0$ we can couple T_1^A and T_1^B by setting

$$T_1^A \stackrel{d}{=} \left\lceil \frac{\ln U_1}{\ln(1 - r_k + d_{k,N})} \right\rceil, \quad T_1^B \stackrel{d}{=} -\frac{\ln U_1}{r_k}$$

for $U_1 \sim \text{Unif}([0, 1])$, since T_1^B is $\text{Exp}(r_{k,k+1} + r_{k,k-1})$ distributed and T_1^A is $\text{Geom}(r_{k,k+1} + r_{k,k-1} + d_{k,N})$ distributed. Note that $T_1^A \geq T_1^B$ almost surely. The coupling holds due to a) if

$$\mathbb{P}(T_2^B - T_1^B < T_1^A - T_1^B) = O(N^{1-2b+2\varepsilon}). \quad (3.49)$$

Furthermore observe

$$T_1^A - T_1^B \geq \ln U_1 \left(\frac{1}{\ln(1 - r_k + d_{k,N})} + \frac{1}{r_k} \right) =: c_k \ln U_1$$

We can upper bound the probability in (3.49) if we assume $T_2^B - T_1^B \sim \text{Exp}(r_{k+1})$, thus we obtain for $E_2 \sim \text{Exp}(r_{k+1})$

$$\begin{aligned} \mathbb{P}(T_2^B - T_1^B < T_1^A - T_1^B) &\leq \mathbb{P}(E_2 \leq c_k \ln U_1) \\ &= 1 - \int_0^1 e^{-r_{k+1} c_k \ln u} du = 1 - \int_0^1 u^{-r_{k+1} c_k} du \\ &= 1 - \frac{1}{r_{k+1} c_k + 1} = 1 - \frac{1}{\frac{r_{k+1}}{\ln(1-r_k+O(e_N))} + \frac{r_{k+1}}{r_k} + 1} \\ &= 1 - \frac{1}{-\frac{r_{k+1}}{r_k + d_{k,N} + O(r_k^2)} + \frac{r_{k+1}}{r_k} + 1} \\ &= 1 - \frac{1}{-\frac{r_{k+1}}{r_k} (1 + O(d_{k,N}/r_k) + O(r_k)) + \frac{r_{k+1}}{r_k} + 1} \\ &= O(d_{k,N}/r_k) + O(r_k) = O(N^{1-2b+2\varepsilon}). \end{aligned}$$

which proves (3.48). Together with Claim 1 this proves the assertion of the lemma. \square

We are now able to complete the proof of Theorem 8a.

Proof of Theorem 8a. Let $(A_m)_{m \in \mathbb{Z}} = (A_m^{(N)})_{m \in \mathbb{Z}}$ be a stationary version of the CASP with parameters N , $\mathcal{L}(\mathcal{W}^{(N)})$ and s_N (where as in the previous statements and proofs we are going to suppress the superscript N in $A_m^{(N)}$). By Corollary 3.3.2 it suffices to analyse $\mathbb{E}[A_0]/N$ in order to obtain the probability of fixation of a single beneficial mutant. Let

$$\mathcal{E} := \mathcal{E}^{(N)} := \{A_{-\lfloor N^{b+\varepsilon} \rfloor} \leq N^{1-b+\varepsilon}, |A_{-j} - B_{-j}| \leq 1, 0 \leq j \leq \lfloor N^{b+\varepsilon} \rfloor\} \quad (3.50)$$

be the event that the (stationary) CASP $(A_m)_{m \in \mathbb{Z}}$ is not unusually big at time $-\lfloor N^{b+\varepsilon} \rfloor$ and can be coupled with a MASP $(B_r)_{r \geq -\lfloor N^{b+\varepsilon} \rfloor}$ for the time between $-N^{b+\varepsilon}$ and 0 such that there the CASP and the MASP differ at most by 1. Due to the Lemmata 3.6.3 and 3.6.9 we can estimate the probability of this event by

$$\mathbb{P}(\mathcal{E}) = (1 - O(N^{-\delta}))(1 - O(\exp(-N^\delta))) = 1 - O(N^{-\delta})$$

and a suitable $\delta > 0$. This yields

$$\frac{\mathbb{E}[A_0]}{N} = \frac{1}{N} \mathbb{E}[A_0 | \mathcal{E}] \mathbb{P}(\mathcal{E}) + \frac{1}{N} \mathbb{E}[A_0 | \mathcal{E}^c] \mathbb{P}(\mathcal{E}^c) \quad (3.51)$$

We analyse the two expectations above separately, the first one will give us the desired Haldane formula, whereas the second is an error term of order $o(s_N)$. By Lemma 3.6.8 we get that with

$$B_{\text{eq}}^{(N)} \stackrel{d}{=} \text{Bin}\left(N, \frac{2s_N}{2s_N + \rho^2}\right) \text{ conditioned to be strictly positive,} \quad (3.52)$$

$$\begin{aligned} \frac{1}{N} \mathbb{E}[A_0 | \mathcal{E}] \mathbb{P}(\mathcal{E}) &= \frac{1}{N} \sum_{j=1}^N j \mathbb{P}(A_0 = j | \mathcal{E}) (1 - O(N^{-\delta})) \\ &= \frac{1}{N} \sum_{j=1}^N j \mathbb{P}\left(B_{\text{eq}}^{(N)} = j\right) (1 - O(N^{-\delta})) \\ &= \frac{1}{N} \frac{N 2s_N}{2s_N + \rho^2} (1 - O(N^{-\delta})) = \frac{2s_N}{\rho^2} (1 + o(s_N)). \end{aligned}$$

It remains to bound the second expectation on the r.h.s. of (3.51), with the worst case being $A_{-\lfloor N^{b+\varepsilon} \rfloor} = N$. Then using the second part of Corollary 3.6.5 gives us

$$\frac{1}{N} \mathbb{E}[A_0 | \mathcal{E}^c] \mathbb{P}(\mathcal{E}^c) = O\left(\frac{N^{1-b+\varepsilon}}{N} N^{-\delta}\right) = O(N^{-b+\varepsilon-\delta}) = o(s_N),$$

since $\varepsilon > 0$ can be chosen small enough such that $\delta > \varepsilon$. This finishes the proof of Theorem 8. \square

Corollary 3.6.10. *Let $A_{\text{eq}}^{(N)}$ have the stationary distribution of the CASP with parameters N , $\mathcal{L}(\mathcal{W}^{(N)})$ and s_N . Then, with $p_N := \frac{s_N}{\rho^2/2 + s_N}$, $\mu_N := N p_N$ and $\sigma_N^2 = N p_N (1 - p_N)$, the sequence of random variables $(A_{\text{eq}}^{(N)} - \mu_N) / \sigma_N$ converges as $N \rightarrow \infty$ in distribution to a standard normal random variable.*

Proof. In the previous proof we have worked, for a stationary CASP $(A_m^{(N)})_{m \in \mathbb{Z}}$, with a decomposition of $\mathbb{E}[A_0^{(N)}]$ according to the events \mathcal{E} and \mathcal{E}^c , with \mathcal{E} defined in (3.50). We now use the same decomposition for the distribution of $(A_0^{(N)} - \mu_N)/\sigma_N$ and obtain for any $f \in C_b(\mathbb{R})$ with the same line of reasoning as in the previous proof and with $B_{\text{eq}}^{(N)}$ as in (3.52)

$$\lim_{N \rightarrow \infty} \mathbb{E}[f((A_0^{(N)} - \mu_N)/\sigma_N)] = \lim_{N \rightarrow \infty} \mathbb{E}[f((B_{\text{eq}}^{(N)} - \mu_N)/\sigma_N)] = \mathbb{E}[f(Z)],$$

where Z is a standard normal random variable. \square

Remark 3.6.11. *Using the technique of [Kur70] it is not difficult to show that $\left(A_{\lfloor rs_N^{-1} \rfloor}^{(N)} / \mu_N\right)_{r \geq 0}$ converges in distribution as $N \rightarrow \infty$ uniformly on compact time intervals to the solution of a dynamical system whose stable fixed point is 1. One might then also ask about the asymptotic fluctuations of the process $A^{(N)}$. Although available results in the literature (like [Kur81, Theorem 8.2] or [EK86, Theorem 11.3.2]) do not directly cover our situation (because e.g. of boundedness assumptions required there), the coupling between $A^{(N)}$ and $B^{(N)}$ analysed above seems a promising tool to obtain a Ornstein-Uhlenbeck functional limit theorem for the fluctuations of $A^{(N)}$, which in view of Corollary 3.6.10 should include also time infinity. Let us mention in this context [Cor17], which contains a fluctuation result (including time infinity) for the Moran frequency process under strong selection and two-way mutation.*

3.7 A concentration result for the equilibrium distribution of the CASP. Proof of Theorem 8

Let $A^{(N)} = (A_m^{(N)})_{m \geq 0}$ be the Cannings ancestral selection process (CASP) as defined in Sec. 3.2.3. We will show in the present section that under the assumptions (3.3), (3.10) and (3.14) the expectation of the equilibrium state $A_{\text{eq}}^{(N)}$ of $A^{(N)}$ satisfies the asymptotics

$$\mathbb{E} \left[A_{\text{eq}}^{(N)} \right] = \frac{2}{\rho^2} s_N N (1 + o(1)). \quad (3.53)$$

The proof of Theorem 8b is then immediate from Corollary 3.3.2.

Let us describe here the strategy of our proof. We will show that the distribution of $A_{\text{eq}}^{(N)}$ is sufficiently concentrated around the “center” $\frac{2}{\rho^2} s_N N$ as $N \rightarrow \infty$. Throughout, we will fix a sequence (h_N) obeying (3.13) such that (3.14) is satisfied. As in the previous section we will switch to b_N defined by (3.25). The assumption (3.3), which is now the standing one, thus translates into

$$\frac{1}{2} + \eta \leq b_N \leq 1 - \eta.$$

Frequently we will suppress the subscript N in b_N , thus denoting the sequence $s_N N$ simply by N^{1-b} . We will show in the subsequent lemmata that the CASP $A^{(N)}$ needs only a relatively short time to enter a small box around $\frac{2}{\rho^2} s_N N$, compared to the time it spends in this box. The former assertion is provided by Lemmata 3.7.4 and 3.7.5. The behaviour of $A^{(N)}$ near the center is controlled by Proposition 3.7.3. This is prepared by Lemmata 3.7.1 and 3.7.2 which bound the probability of jumps of absolute size larger than h_N near the center. This allows to couple the process $\mathcal{A} := A^{(N)}$ with a process \mathcal{A}^u which in a certain box close to the center replaces all the jumps of \mathcal{A} of size $2, \dots, h_N$ by an upward jump of size h_N and is constructed

such that it stochastically dominates \mathcal{A} on an event of high probability. See Figures 3.2 and 3.3 for illustrations of this construction; the definition of \mathcal{A}^u is given in the proof of Proposition 3.7.3.

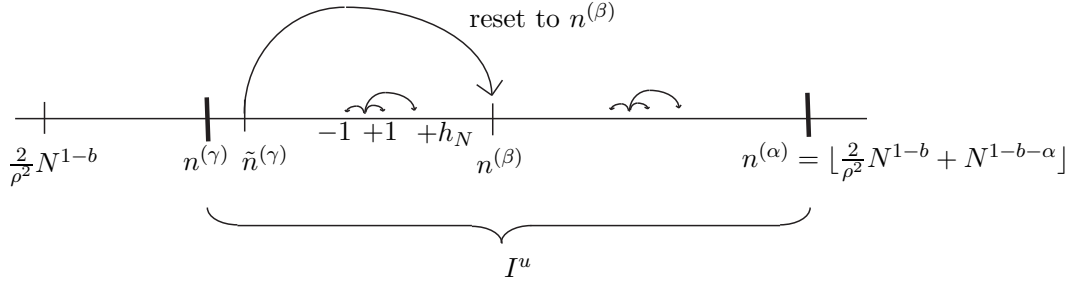


FIGURE 3.2: A sketch of the transition dynamics of the “upper stochastic dominant” \mathcal{A}^u of the CASP \mathcal{A} : In the box $I^u = [\tilde{n}^{(\gamma)}, n^{(\alpha)}]$, the process \mathcal{A}^u performs jumps of size ± 1 whenever \mathcal{A} makes a jump ± 1 , and makes a jump $+h_N$ whenever \mathcal{A} makes a jump $+2, \dots, +h_N$. Whenever \mathcal{A}^u reaches $\tilde{n}^{(\gamma)}$, it is reset to its starting point $n^{(\beta)}$. The quantities $n^{(\beta)}$ and $\tilde{n}^{(\gamma)}$ as well as the transition probabilities of \mathcal{A}^u are defined in the proof of Proposition 3.7.3.

The following lemma controls the probability of large upward jumps of \mathcal{A} near the center, using the construction of the branching step of the CASP described in Section 3.2.3.

Lemma 3.7.1 (Probability for large jumps upwards).

Let $k = \lceil \kappa N^{1-b} \rceil$ for some $\kappa > 0$, then

$$\mathbb{P}(A_{m+1} \geq A_m + h_N | A_m = k) = O((N^{1-2b})^{h_N}) \text{ as } N \rightarrow \infty.$$

Proof. We want to estimate

$$\mathbb{P}(A_{m+1} \geq A_m + h_N | A_m = k) \leq \mathbb{P}\left(\sum_{i=1}^k G^{(i)} \geq k + h_N\right),$$

for independent $\text{Geom}(p)$ with $p = 1 - s_N$ distributed random variables $G^{(i)}$, $i \geq 1$. With $k' = k + h_N$, $S_{k'}$ a $\text{Bin}(k', p)$ -distributed random variable, and $a = \frac{k}{k'}$, we can estimate the r.h.s. from above by

$$\mathbb{P}(S_{k'} \leq k) = \mathbb{P}(S_{k'} \leq ak'),$$

Using the Chernoff bound for binomials we can estimate

$$\mathbb{P}(S_{k'} \leq ak') \leq \exp(-k'H(a)), \tag{3.54}$$

with rate function $H(a) = a \ln\left(\frac{a}{p}\right) + (1-a) \ln\left(\frac{1-a}{1-p}\right)$. Inserting our parameters yields

$$\begin{aligned} H(a) &= \frac{k}{k+h_N} \ln\left(\frac{k}{k+h_N} \frac{1}{1-s_N}\right) + \frac{h_N}{h_N+k} \ln\left(\frac{h_N}{k+h_N} \frac{1}{s_N}\right) \\ &= \frac{\kappa N^{1-b}}{\kappa N^{1-b} + o(N^{1-b})} \ln\left(1 - \frac{h_N}{\kappa N^{1-b}}(1+o(1))\right) \\ &\quad + \frac{h_N}{\kappa N^{1-b} + o(N^{1-b})} \ln\left(\frac{N^b h_N}{\kappa N^{1-b} + o(N^{1-b})}\right) \\ &= \frac{h_N}{\kappa N^{1-b}} \ln\left(\frac{1}{\kappa} N^{2b-1} h_N\right) (1+o(1)) - \frac{1}{\kappa} N^{b-1} h_N (1+o(1)). \end{aligned}$$

The dominating term above is $\frac{1}{\kappa} N^{b-1} h_N \ln(N^{2b-1})$ and plugging this back into (3.54) one obtains

$$\begin{aligned} \mathbb{P}(S_{n'} \leq an') &\leq \exp\left(-\kappa N^{1-b} \frac{1}{\kappa} N^{b-1} h_N \ln(N^{2b-1})(1+o(1))\right) \\ &= \exp(-h_N \ln(N^{2b-1}))(1+o(1)) = (N^{1-2b})^{h_N} (1+o(1)). \end{aligned}$$

□

Next, we set out to bound the probability for downward jumps of size at least h_N near the center. In view of the construction of the coalescence step described in Section 3.2.3 this is settled by the following lemma.

Lemma 3.7.2 (Probability for large jumps downwards).

Let $\mathscr{W}^{(N)} = (W_1^{(N)}, \dots, W_N^{(N)})$ be as in Theorem 3.5b, let $\kappa > 0$ and $k^{(N)} \leq \kappa N^{1-b}$. For $N \in \mathbb{N}$ sort $k^{(N)}$ balls independently into N boxes, such that the ball with number ν is sorted into box i with probability $W_i^{(N)}$, $i \in [N]$, $\nu \in [k^{(N)}]$. Then the probability that no more than $k^{(N)} - h_N$ boxes are occupied is

$$O\left(h_N^4 N^{1-2b}\right)^{h_N} \text{ as } N \rightarrow \infty.$$

Proof. We will suppress the superscript (N) and write $k := k^{(N)}$, $\mathscr{W} := \mathscr{W}^{(N)}$. For $h := h_N$ let p_h be the probability of the event that no more than $k - h$ boxes are occupied. This is equal to the event that at least h collisions occur, where we think of the balls with numbers $1, \dots, k$ being subsequently sorted into the boxes and say that the ball with number ν produces a collision if it lands in an already occupied box. In the following we record the occupation numbers of (only) those boxes that receive more than one ball. These are of the form $\beta = (\beta_1, \dots, \beta_\ell) \in \{2, \dots, h+1\}^\ell$ with $\ell \in \{1, \dots, h\}$ and $\beta_1 + \dots + \beta_\ell - \ell = h$. For a given β of this form, and ℓ given boxes with $|\beta| := \beta_1 + \dots + \beta_\ell$, assume that β_1 balls are sorted into the first box, β_2 balls into the second box, etc, and the remaining $k - |\beta|$ balls are sorted into arbitrary boxes (so that, as required, the number of occupied boxes is at most $\ell + k - |\beta| = k - h$). Given the weights W_1, \dots, W_N , the probability to sort the first β_1 balls into box 1, the following β_2 balls into box 2, \dots , and finally β_ℓ balls into box ℓ is

$$\prod_{i=1}^{\ell} W_i^{\beta_i}.$$

There are $\frac{N!}{(N-\ell)!}$ many possibilities to choose ℓ different boxes out of N . Furthermore, there are $\binom{k}{\beta_1, \dots, \beta_\ell, k-|\beta|}$ many possibilities to choose $|\beta|$ many balls out of k balls and sort these balls into ℓ boxes, such that β_i balls are sorted into box i .

Due to exchangeability of the weights W_1, \dots, W_N we get

$$p_h \leq \sum_{\beta \in \mathcal{B}} \mathbb{E} \left[\prod_{i=1}^{\ell(\beta)} W_i^{\beta_i} \right] \frac{N!}{(N-\ell(\beta))!} \binom{k}{\beta_1, \dots, \beta_\ell, k-|\beta|}, \quad (3.55)$$

where

$$\mathcal{B} := \bigcup_{\ell \in \{1, \dots, h\}} \{ \beta = (\beta_1, \dots, \beta_\ell) : \beta_1 + \dots + \beta_\ell - \ell = h \text{ and } \beta_1 \geq \beta_2 \geq \dots \geq \beta_\ell \}$$

and $\ell = \ell(\beta)$ denotes the length of the vector $\beta \in \mathcal{B}$.

To obtain an upper bound of the r.h.s. of (3.55) we estimate the moments $\mathbb{E} \left[\prod_{i=1}^{\ell} W_i^{\beta_i} \right]$. Since (W_1, W_2, \dots, W_N) are negatively associated, we can use the property 2 in [JDP83] of negatively associated random variables, which reads

$$\mathbb{E} \left[\prod_{i=1}^{\ell} W_i^{\beta_i} \right] \leq \prod_{i=1}^{\ell} \mathbb{E} \left[W_i^{\beta_i} \right]$$

Applying Jensen's inequality we can estimate for $\beta_1 \geq \beta_2$

$$\begin{aligned} \mathbb{E} \left[W_1^{\beta_1} \right] \mathbb{E} \left[W_2^{\beta_2} \right] &\leq \mathbb{E} \left[W_1^{\beta_1} \right] \mathbb{E} \left[W_2^{\beta_1} \right]^{\frac{\beta_2}{\beta_1}} = \mathbb{E} \left[W_1^{\beta_1} \right]^{1 + \frac{\beta_2}{\beta_1}} \\ &\leq \mathbb{E} \left[W_1^{\beta_1 + \beta_2} \right]. \end{aligned}$$

Hence,

$$\prod_{i=1}^{\ell} \mathbb{E} \left[W_i^{\beta_i} \right] \leq \mathbb{E} \left[W_1^{|\beta|} \right].$$

We will now analyse the quantities

$$a_\beta := \mathbb{E} \left[W_1^{|\beta|} \right] \frac{N!}{(N-\ell(\beta))!}, \quad \beta \in \mathcal{B}. \quad (3.56)$$

For N sufficiently large, and any two occupation vectors $\beta, \gamma \in \mathcal{B}$ with $\ell(\beta) = \ell(\gamma) + 1$ and $|\beta| = |\gamma| + 1$ we obtain from (3.14)

$$\frac{a_\beta}{a_\gamma} = \frac{\mathbb{E} \left[W_1^{|\gamma|+1} \right] \frac{N!}{(N-\ell(\gamma)+1)!}}{\mathbb{E} \left[W_1^{|\gamma|} \right] \frac{N!}{(N-\ell(\gamma))!}} \leq \frac{h_N}{N} (N - \ell(\gamma)) \leq h_N. \quad (3.57)$$

For the occupation vector $\tilde{\beta} := (1 + h_N) \in \mathcal{B}$ (with length $\ell(\tilde{\beta}) = 1$ and $|\tilde{\beta}| = 1 + h_N$) we have, again because of (3.14),

$$a_{\tilde{\beta}} = \mathbb{E} \left[W_1^{h_N+1} \right] N = O \left(\left(\frac{h_N}{N} \right)^{h_N} \right). \quad (3.58)$$

Combining (3.57) and (3.58) we obtain for all $\beta \in \mathcal{B}$ the estimate

$$a_\beta \leq O\left(N^{-h_N}(h_N)^{2h_N}\right). \quad (3.59)$$

For the rightmost term in (3.55) we have the estimate

$$\binom{k}{\beta_1, \dots, \beta_\ell, k - |\beta|} \leq \binom{k}{2, \dots, 2, k - 2h_N} \leq (\kappa N^{1-b})^{2h_N} \quad (3.60)$$

and the number of occupation vectors β appearing in the sum in (3.55) (i.e. the cardinality of \mathcal{B}) can be estimated from above by $(h_N + 1)^{h_N}$. Hence we obtain from (3.55), (3.56), (3.59) and (3.60),

$$\begin{aligned} p_{h_N} &= O\left((\kappa N^{1-b})^{2h_N} N^{-h_N} (h_N^{2h_N} (h_N + 1)^{h_N})\right) \\ &= O\left((N^{1-2b} h_N^4)^{h_N}\right). \end{aligned}$$

□

Building on the previous two lemmata, the next result shows that the CASP does not leave the central region up to any polynomially long time with high probability.

Proposition 3.7.3 (CASP stays near the center for a long time).

Consider α, β with $0 < \alpha < \beta < \frac{2b-1}{3}$, let

$$I := \left[\left[\frac{2}{\rho^2} N^{1-b} - N^{1-b-\alpha} \right], \left[\frac{2}{\rho^2} N^{1-b} + N^{1-b-\alpha} \right] \right]$$

and define $\tau_I := \inf\{m \geq 0 : A_m \notin I\}$. Then for all $\theta > 0$ and all $\varepsilon > 0$

$$\mathbb{P}\left(\tau_I \leq N^\theta \mid A_0 = k\right) = O\left((N^{1-2b+\varepsilon})^{h_N}\right) \quad \text{as } N \rightarrow \infty \quad (3.62)$$

uniformly in $k \in \left[\left\lceil \frac{2}{\rho^2} N^{1-b} - N^{1-b-\beta} \right\rceil, \left\lfloor \frac{2}{\rho^2} N^{1-b} + N^{1-b-\beta} \right\rfloor\right]$.

Proof. To show the above claim we bound stochastically \mathcal{A} from above and from below by simpler processes, which in certain boxes close to the center of attraction of \mathcal{A} follow essentially a time changed random walk dynamics with constant drift. In the first part of the proof we will construct a time-changed Markov chain \mathcal{A}^u that dominates $\mathcal{A} = (A_m)_{m \geq 0}$ from above for a sufficiently long time. This construction will rely on Lemma 3.7.1. In this first part we will give all details; in the second part of the proof we will indicate how an analogous construction can be carried out “from below”, then making use of 3.7.2.

1. Let $\gamma \in (\beta, \frac{2b-1}{3})$ and write, in accordance with Figure 3.2,

$$n^{(\zeta)} := \left\lceil \frac{2N^{1-b}}{\rho^2} + N^{1-b-\zeta} \right\rceil, \quad \zeta \in \{\alpha, \beta, \gamma\}.$$

Observing that $n^{(\gamma)} < n^{(\beta)} < n^{(\alpha)}$, we consider the box

$$I^u := \left[n^{(\gamma)}, n^{(\alpha)} \right],$$

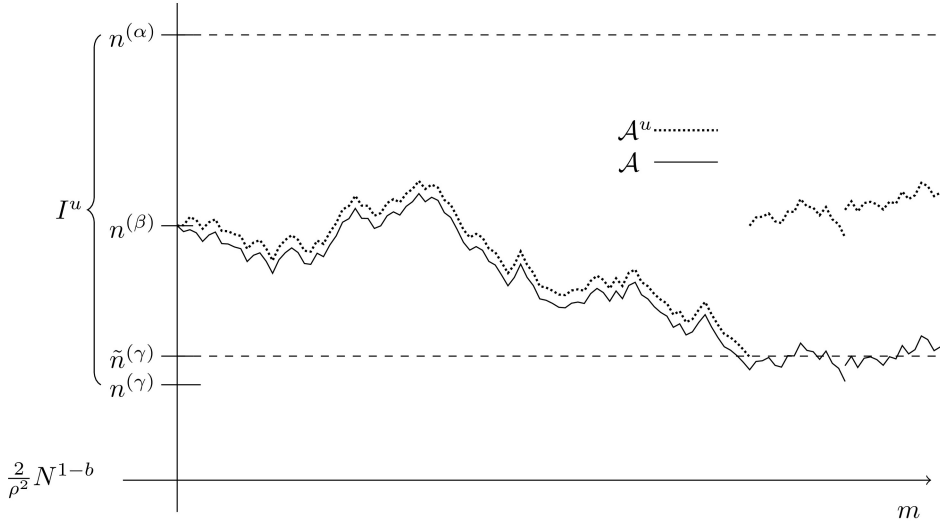


FIGURE 3.3: An example of a realisation of the processes \mathcal{A} and \mathcal{A}^u , displaying that \mathcal{A}^u dominates \mathcal{A} as long as it is below the level $n^{(\alpha)}$. Note that \mathcal{A}^u is reset to $n^{(\beta)}$ whenever it hits the level $\tilde{n}^{(\gamma)}$; see also Figure 3.2.

and take $n^{(\beta)}$ as the starting point of both \mathcal{A} and \mathcal{A}^u . The dynamics of \mathcal{A}^u will be such that \mathcal{A}^u is re-set to its starting point $n^{(\beta)}$ as soon as it hits the level

$$\tilde{n}^{(\gamma)} := n^{(\gamma)} + h_N.$$

Consider the following Markov chain $\bar{\mathcal{A}}^u$. We decree that within the box $(\tilde{n}^{(\gamma)}, n^{(\alpha)})$ the process $\bar{\mathcal{A}}^u$ makes only jumps $-1, +1$ and $+h_N$. Here, the probabilities for jumps $-1, +1$ of $\bar{\mathcal{A}}^u$ from an arbitrary state in $(\tilde{n}^{(\gamma)}, n^{(\alpha)})$ are set equal to the probabilities for jumps $-1, +1$ of \mathcal{A} from the state $n^{(\gamma)}$, and the probability for a jump $+h_N$ of $\bar{\mathcal{A}}^u$ from an arbitrary state in $(\tilde{n}^{(\gamma)}, n^{(\alpha)})$ is set equal to the probability of a jump of \mathcal{A} from the state $n^{(\gamma)}$ that has an absolute size larger than 1. More formally, we define

$$c_N^{(\gamma)} = \mathbb{P}(|A_{m+1} - A_m| > 0 | A_m = n^{(\gamma)})$$

and observe that

$$c_N^{(\gamma)} = \mathbb{P}(|A_{m+1} - A_m| > 0 | A_m = n^{(\gamma)}) = s_N n^{(\gamma)} + \binom{n^{(\gamma)}}{2} \frac{\rho^2}{N} (1 + o(1)).$$

We define the jump probabilities as follows:

$$\mathbb{P}(\bar{A}_i^u = k + 1 | \bar{A}_{i-1}^u = k) = \frac{1}{c_N^{(\gamma)}} s_N n^{(\gamma)}, \quad k > \tilde{n}^{(\gamma)},$$

$$\mathbb{P}(\bar{A}_i^u = k + h_N | \bar{A}_{i-1}^u = k) = 1 - \frac{1}{c_N^{(\gamma)}} \left(s_N n^{(\gamma)} + \binom{n^{(\gamma)}}{2} \frac{\rho^2}{N} \right), \quad k > \tilde{n}^{(\gamma)}$$

$$\mathbb{P}(\bar{A}_i^u = k - 1 | \bar{A}_{i-1}^u = k) = \frac{1}{c_N^{(\gamma)}} \binom{n^{(\gamma)}}{2} \frac{\rho^2}{N}, \quad k > \tilde{n}^{(\gamma)},$$

$$\mathbb{P}(\bar{A}_i^u = n^{(\beta)} | \bar{A}_{i-1}^u = \tilde{n}^{(\gamma)}) = 1.$$

Note that

$$\begin{aligned} \frac{1}{c_N^{(\gamma)}} s_N n^{(\gamma)} &= \frac{1}{2} \left(1 - \frac{\rho^2}{4} N^{-\gamma} \right) (1 + O(s_N n^{(\gamma)})), \\ \frac{1}{c_N^{(\gamma)}} \binom{n^{(\gamma)}}{2} \frac{\rho^2}{N} &= \frac{1}{2} \left(1 + \frac{\rho^2}{4} N^{-\gamma} \right) (1 + O(s_N n^{(\gamma)})), \\ 1 - \frac{1}{c_N^{(\gamma)}} \left(s_N n^{(\gamma)} + \binom{n^{(\gamma)}}{2} \frac{\rho^2}{N} \right) &= O(N^{1-2b}), \end{aligned}$$

which results in a small downwards drift of \bar{A}^u in I^u . The process \mathcal{A}^u is defined as follows. Denote by τ_i the time of the i -th non-trivial jump (that is a jump of size $\neq 0$) of \mathcal{A} for $i \geq 1$ and let $\tau_0 = 0$. We set for $\tau_{i-1} \leq m \leq \tau_i - 1$, with $i \geq 1$,

$$A_m^u = \bar{A}_{i-1}^u.$$

Putting $T_N := \lceil N^\theta \rceil$, we now consider the events

$$E_N := \{\mathcal{A}^u \text{ has not left } I^u \text{ by time } T_N\}$$

and

$$F_N := \{\mathcal{A} \text{ has not performed jumps of absolute size larger than } h_N \text{ by time } T_N\}.$$

We can now couple \mathcal{A} and \mathcal{A}^u such that on the event $E_N \cap F_N$ and for all $m \leq T_N$ we have $A_m \leq A_m^u$. In order to show that the probability of the event $\{\mathcal{A} \text{ reaches } n^{(\alpha)} \text{ before time } T_N\}$ is bounded by the r.h.s. of (3.62) it thus suffices to show that

$$\mathbb{P}(E_N^c) = O((N^{1-2b+\varepsilon})^{h_N}) \quad \text{and} \quad \mathbb{P}(F_N^c) = O((N^{1-2b+\varepsilon})^{h_N}) \quad \text{as } N \rightarrow \infty. \quad (3.63)$$

From Lemmata 3.7.1 and 3.7.2 it is obvious that

$$\mathbb{P}(F_N^c) = O(T_N (h_N^4 N^{1-2b})^{h_N}). \quad (3.64)$$

We claim that there exists a $\delta > 0$ such that

$$\mathbb{P}(E_N^c) = O(\exp(-N^\delta)). \quad (3.65)$$

Write p_{hit} for the probability that \mathcal{A}^u , when started from $n^{(\beta)}$, hits (or crosses) $n^{(\alpha)}$ before it hits $\tilde{n}^{(\gamma)}$. The jump size of the process \mathcal{A}^u is at least -1 at each generation. Therefore, at least $N^{1-b-\beta}$ generations are necessary to reach the level $n^{(\gamma)}$ when starting from level $n^{(\beta)}$. Thus, with the θ given in the Proposition, within N^θ generations the process \mathcal{A}^u makes at most $N^\theta / N^{1-b-\beta}$ transitions from $n^{(\beta)}$ to $n^{(\gamma)}$.

The probability that \mathcal{A}^u crosses $n^{(\alpha)}$ within a single excursion from $n^{(\beta)}$ that reaches $\tilde{n}^{(\gamma)}$ is obviously upper bounded by the probability that the process $\tilde{A}_m = \sum_{i=1}^m X_i$, $m = 0, 1, \dots$, crosses the level $n^{(\alpha)}$, where the X_i are i.i.d. and distributed as the jump sizes of the process \bar{A}^u in a state $x > \tilde{n}^\gamma$; thus, other than the process \bar{A}^u the process \tilde{A} is not reset to $n^{(\beta)}$ at an attempt to cross $\tilde{n}^{(\gamma)}$. For $\lambda_N = N^{-2\gamma}$ one has

$$\mathbb{E}[\exp(\lambda_N X_1)] = 1 - \frac{\rho^2}{4} N^{-3\gamma} + O(N^{1-2b}) + o(N^{-3\gamma}) < 1,$$

for N large enough, since $\gamma < \frac{2b-1}{3}$. Therefore the process $Y_m = \exp(\lambda_N \sum_{i=1}^m X_i)$

is a supermartingale. Define $\tilde{\tau} = \inf\{m \geq 0 : \tilde{A}_m \geq n^{(\alpha)}\}$, then by the Martingale Stopping Theorem

$$n^{(\beta)} = \mathbb{E}[Y_0] \geq \mathbb{E}[Y_{m \wedge \tilde{\tau}}] \geq \mathbb{E}[Y_{\tilde{\tau}} \mathbb{1}_{\{\tilde{\tau} < m\}}] \geq \exp(\lambda_N n^{(\alpha)}) \mathbb{P}(\tilde{\tau} < m),$$

which yields the Cramér-Lundberg bound $p_{\text{hit}} \leq n^{(\beta)} \exp(-\lambda_N n^{(\alpha)})$. Thus, we can estimate

$$\mathbb{P}(E_N) \geq (1 - p_{\text{hit}})^{N^{\theta-1+b+\beta}},$$

which proves (3.65). Clearly, (3.65) and (3.64) imply (3.63), which completes the first part of the proof.

2. It remains to prove also the “lower part” of (3.62), i.e. to control the time it takes \mathcal{A} to leave I in downwards direction. We argue similarly by defining a process \mathcal{A}^ℓ which bounds \mathcal{A} from below in a box I^ℓ correspondingly chosen to the box I^u . This process is again a Markov chain that makes jumps of size 1 and -1 and (rarely) jumps of size $-h_N$ and whose drift coincides with that of \mathcal{A} at the upper boundary of the box I^ℓ . Due to Lemma 3.7.2 jumps downwards at least of size h_N occur with exponentially small probability and hence, these jumps can be ignored in the time frame of interest and \mathcal{A}^ℓ is stochastically dominated by \mathcal{A} with sufficiently high probability. \square

We now show that the time to reach the center from state N is with high probability no larger than polynomial.

Lemma 3.7.4 (Coming down from N).

Let $0 < \varepsilon < \frac{2b-1}{3}$, $B := [\frac{2}{\rho^2} N^{1-b} - 2N^{1-b-\varepsilon}, \frac{2}{\rho^2} N^{1-b} + 2N^{1-b-\varepsilon}]$ and $\tau_B := \inf\{m \geq 0 : A_m \in B\}$. Then for any $\varepsilon < \varepsilon' < \frac{2b-1}{3}$ and $0 < \delta \leq \varepsilon' - \varepsilon$

$$\mathbb{P}(\tau_B \geq N^{2b+2\varepsilon'} \mid A_0 = N) = O(\exp(-N^\delta)).$$

Proof. The proof will be divided in three steps.

1. The coalescence probabilities of the Wright-Fisher model are the smallest in our class of Cannings models with selection, see Lemma 3.6.4. Therefore, the stopping time τ_B for the CASP is stochastically dominated from above by the corresponding stopping time in a Wright-Fisher model with selection. Consequently, we assume in the following that $\mathscr{W} = (1/N, \dots, 1/N)$ and $\rho^2 = 1$.

2. We analyse the drift of $(A_m)_{m \geq 0}$ in each point y in $B' = [2N^{1-b} + 2N^{1-b-\varepsilon'}, N]$. We will show in part 3 of the proof that

$$y - \mathbb{E}[A_1 | A_0 = y] \geq 2N^{1-2b-\varepsilon'}, \quad \text{for all } y \in B'. \quad (3.66)$$

The estimate (3.66) on the drift of \mathcal{A} in B' yields that for $m_0 = \lceil N^{2b+\varepsilon'} \rceil$ it holds $\mathbb{E}[A_{m_0}] \leq 2N^{1-b} + 2N^{1-b-\varepsilon}$, since by Proposition 3.7.3 after entering the box B the process \mathcal{A} does not leave the box up to any polynomially long time except on an event with probability $O((N^{1-2b+\varepsilon})^{h_N})$. Applying Markov's inequality yields for

$$m_0 = \lceil N^{2b+\varepsilon'} \rceil$$

$$\begin{aligned} \mathbb{P}(A_{m_0} \geq 2N^{1-b} + 2N^{1-b-\varepsilon}) &\leq \frac{\mathbb{E}[A_{m_0}]}{2N^{1-b} + 2N^{1-b-\varepsilon}} \leq \frac{2N^{1-b} + 2N^{1-b-\varepsilon'}}{2N^{1-b} + 2N^{1-b-\varepsilon}} \\ &= 1 + \frac{N^{1-b-\varepsilon'} - N^{1-b-\varepsilon}}{N^{1-b} + N^{1-b-\varepsilon}} = 1 + \frac{N^{-\varepsilon'} - N^{-\varepsilon}}{1 + N^{-\varepsilon}} \\ &= 1 - N^{-\varepsilon}(1 + N^{-(\varepsilon'-\varepsilon)})(1 + O(N^{-\varepsilon})). \end{aligned}$$

Hence, after time $\lceil N^{2b+\varepsilon'} \rceil$ the process started in N is with a probability of at least $1 - N^{-\varepsilon}(1 + o(1))$ in B . If the process \mathcal{A} did not enter B until time $\lceil N^{2b+\varepsilon'} \rceil$, in the worst case the process is still in the state N . Therefore, recalling that $\delta \leq \varepsilon' - \varepsilon$, the probability that the process is after time $\lceil N^{2b+2\varepsilon'} \rceil$ still above B can be estimated from above via

$$\begin{aligned} \mathbb{P}(A_{\lceil N^{2b+2\varepsilon'} \rceil} \geq 2N^{1-b} + 2N^{1-b-\varepsilon}) &\leq \mathbb{P}(A_{m_0} \geq 2N^{1-b} + 2N^{1-b-\varepsilon})^{N^{\varepsilon'}} \leq (1 - N^{-\varepsilon})^{N^{\varepsilon'}} \\ &= O(\exp(-N^\delta)). \end{aligned}$$

3. It remains to show (3.66). Recalling the ‘‘balls in boxes’’ description of the one-step transition probability of the CASP as described in Sec. 3.2.3, let χ_i be the indicator of the event that the i -th box is occupied by at least one ball. We can rewrite

$$\begin{aligned} y - \mathbb{E}[A_1 | A_0 = y] &= y - \mathbb{E}\left[\sum_{i=1}^N \chi_i | A_0 = y\right] \\ &= y - N \left(1 - \mathbb{E}\left[\left(1 - \frac{1}{N}\right)^H\right]\right) \end{aligned} \quad (3.69)$$

with $H = \sum_{i=1}^y G^{(i)}$ with $G^{(1)}, \dots, G^{(y)}$ independent and $G^{(i)} \sim \text{Geom}(1 - s_N)$. The expectation in (3.69) is the generating function of a negative binomial distribution with parameters y and $(1 - s_N)$ evaluated at $1 - \frac{1}{N}$, which allows to continue (3.69) as

$$\begin{aligned} &= y - N \left(1 - \left[\frac{(1 - s_N)(1 - \frac{1}{N})}{1 - s_N(1 - \frac{1}{N})}\right]^y\right) \\ &= y - N \left(1 - \left[1 - \frac{1}{N} \frac{1}{1 - s_N + \frac{s_N}{N}}\right]^y\right) \end{aligned} \quad (3.70)$$

Using Taylor expansion for the function $x \mapsto (1 - x)^y$ around 0 with the remainder in Lagrange form yields

$$(1 - x)^y = 1 - yx + \binom{y}{2}x^2 - \binom{y}{3}x^3(1 - \xi)^{y-3} \geq 1 - yx + \binom{y}{2}x^2 - \binom{y}{3}x^3 \quad (3.71)$$

for some $\xi \in [0, x]$. Abbreviating $u := \frac{1}{1 - s_N + \frac{s_N}{N}}$ in (3.69) and using (3.71) yields

$$\begin{aligned} (3.70) &\geq y - N \left(1 - \left[1 - y\frac{u}{N} + \binom{u}{2}\frac{u^2}{N^2} - \binom{y}{3}\frac{u^3}{N^3}\right]\right) \\ &= y - uy + \frac{u^2}{N} \binom{y}{2} - \frac{u^3}{N^2} \binom{y}{3} =: h(y). \end{aligned}$$

In order to show that the polynomial $h(y)$ is bounded away from 0 as claimed in (3.66) we check that h is positive at the lower boundary and that the derivative h' is positive on the interval B' . We can factorise h as $h(y) = y\tilde{h}(y)$, with

$$\tilde{h}(y) = 1 - u + \frac{u^2}{2N}(y-1) - \frac{u^3}{6N^2}(y-1)(y-2)$$

It is straightforward to check for $y_0 = 2N^{1-b} + 2N^{1-b-\varepsilon'}$ that $\tilde{h}(y_0) > 0$. Thus it suffices to show that \tilde{h}' is strictly positive on B' . We have

$$\tilde{h}'(y) = -\frac{u^3}{3N^2}y + \frac{1}{2}\frac{u^3}{N^2} + \frac{u^2}{2N}.$$

Hence $\tilde{h}'(y) > 0$ is implied by the inequality

$$0 \leq y \leq \frac{3}{2}\frac{N}{u} + \frac{1}{2}$$

which is fulfilled for all $y \in B'$. Hence the drift $\mathbb{E}[A_1|A_0 = y] - y$ is negative for all states $y \in B'$ with minimal absolute value bounded from below by

$$h(y_0) \geq 2N^{1-2b-\varepsilon'}(1 + O(N^{-\varepsilon'})).$$

which proves (3.66). □

Similar as Lemma 3.7.4 we show now that the time to reach the box B from below is also no longer than polynomial with high probability.

Lemma 3.7.5 (Going up from 1).

Let B and τ_B as in Lemma 3.7.4. Then for $0 < \varepsilon < \varepsilon' < \frac{1-2b}{3}$ and $\delta \leq \varepsilon' - \varepsilon$.

$$\mathbb{P}\left(\tau_B \geq N^{2b+2\varepsilon'} \mid A_0 = 1\right) = O((N^{1-2b+\varepsilon})^{h_N}).$$

Proof. The strategy of the proof is similar to the proof of Lemma 3.7.4: we estimate the drift in each point y in the box B' from below, with $B' := [1, \frac{2}{\rho^2}N^{1-b} - N^{1-b-\varepsilon'}]$.

1. Assume we have shown for $y \in B'$ and N large enough

$$\mathbb{E}[A_1|A_0 = y] - y \geq y\frac{\rho^2}{4}N^{-b-\varepsilon'}. \quad (3.72)$$

Then

$$\mathbb{E}[A_{m_0}|A_0 = 1] \geq \inf\left\{\left(1 + \frac{\rho^2}{4}N^{-b-\varepsilon'}\right)^{m_0}, \frac{2}{\rho^2}N^{1-b} - N^{1-b-\varepsilon'}\right\},$$

as long as m_0 is at most of polynomial order in N , since due to Proposition 3.7.3 after entering the box B the process A_m does not leave the box up to any polynomially long time with probability $1 - O((N^{1-2b+\varepsilon})^{h_N})$. In particular, for $m_0 = \lceil cN^{b+\varepsilon'} \ln N \rceil$ and some constant $c > 0$, we have $\mathbb{E}[A_{m_0}] \geq \frac{2}{\rho^2}N^{1-b} - N^{1-b-\varepsilon'}$.

Now applying Markov's inequality yields

$$\begin{aligned}
\mathbb{P}(A_{m_0} \leq \frac{2}{\rho^2} N^{1-b} - N^{1-b-\varepsilon}) &= \mathbb{P}(N - A_{m_0} \geq N - \frac{2}{\rho^2} N^{1-b} + N^{1-b-\varepsilon}) \\
&\leq \frac{\mathbb{E}[N - A_{m_0}]}{N - \frac{2}{\rho^2} N^{1-b} + N^{1-b-\varepsilon}} \leq \frac{N - \frac{2}{\rho^2} N^{1-b} + N^{1-b-\varepsilon'}}{N - \frac{2}{\rho^2} N^{1-b} + N^{1-b-\varepsilon}} \\
&= 1 - N^{-b-\varepsilon}(1 + o(1)) \tag{3.73}
\end{aligned}$$

In the worst case at time m_0 the process A is still in state 1. Iterating the argument in (3.73) yields that the probability that after time $\lceil N^{b+\varepsilon'} \rceil m_0$ the process is still below $\frac{2}{\rho^2} N^{1-b} - N^{1-b-\varepsilon}$ is of order $O(\exp(-N^\delta))$, as claimed in the Lemma.

2. Now it remains to show (3.72). For $y \in B'$ we observe that

$$\begin{aligned}
\mathbb{E}[A_1 | A_0 = y] - y &= N(1 - \mathbb{E}[(1 - W_1)^H | A_0 = y]) - y \\
&= N(1 - \mathbb{E}[\mathbb{E}[(1 - W_1)^H | A_0 = y, W_1]]) - y \\
&= N(1 - \mathbb{E}\left[\left(1 - \frac{W_1}{1 - s_N + s_N W_1}\right)^y \middle| A_0 = y\right]) - y \\
&= N \mathbb{E}\left[\frac{y W_1}{1 - s_N + s_N W_1} - \frac{\binom{y}{2} W_1^2}{(1 - s_N + s_N W_1)^2} + O(y^3 W_1^3)\right] - y \\
&\geq N \mathbb{E}\left[\frac{y W_1}{1 - s_N + s_N W_1} - \frac{\binom{y}{2} W_1^2}{(1 - s_N)^2} + O(y^3 W_1^3)\right] - y \tag{3.74}
\end{aligned}$$

We analyse the first summand in the expectation separately, since the denominator and nominator both contain the random variable W_1 .

$$\begin{aligned}
\mathbb{E}\left[y \frac{W_1}{1 - s_N + s_N W_1}\right] &= \mathbb{E}[y W_1 (1 + s_N + O(s_N W_1 + s_N^2))] \\
&= \frac{y}{N} (1 + s_N + O(s_N^2))
\end{aligned}$$

Using that $\mathbb{E}[W_1^3] = O(N^{-3})$ (which follows from the assumptions of Theorem 8), one can continue (3.74) as

$$\begin{aligned}
&= y(1 + s_N + O(s_N^2)) - \binom{y}{2} \frac{\rho^2}{N} \frac{1}{(1 - s_N)^2} - y + O(y N^{-b-1}) + O(y^3 \frac{1}{N^2}) \\
&\geq y \left(1 + s_N + O(s_N^2) - \frac{s_N - \frac{\rho^2}{2} s_N N^{-\varepsilon'}}{(1 - s_N)^2} - 1 + O(N^{-b-1}) + O(\frac{y^2}{N^2})\right) \\
&\geq y \frac{\rho^2}{4} s_N N^{-\varepsilon'},
\end{aligned}$$

where in the first inequality we used that $y \in B'$ is bounded from above. This gives the desired estimate (3.72) for the drift, and completes the proof of the lemma. \square

Completion of the proof of Theorem 8b.

1. For proving (3.53) we will make use of Corollary 3.3.2, and to this purpose derive asymptotic upper and lower bounds on the expectation of $A_{\text{eq}} := A_{\text{eq}}^{(N)}$ via stochastic domination from above and below. Consider a time-stationary version $\mathcal{A}^{\text{stat}} = (A_m^{\text{stat}})_{m \in \mathbb{Z}}$ and a CASP $\mathcal{A} = (A_m)_{m \geq 0}$ that is started in N . We can couple

both processes such that a.s. $A_m \geq A_m^{\text{stat}}$ for all $m \geq 0$. This implies

$$\mathbb{E}[A_{\text{eq}}] = \mathbb{E}[A_m^{\text{stat}}] \leq \mathbb{E}[A_m], \quad m \geq 0.$$

Fix $0 < \alpha < \beta < \frac{2b-1}{3}$, and consider the box $B^\alpha = [\frac{2}{\rho^2}N^{1-b} \pm N^{1-b-\alpha}]$ as well as the (smaller) box $B^\beta = [\frac{2}{\rho^2}N^{1-b} \pm N^{1-b-\beta}]$. Define $\tau_{B^\beta} := \inf\{m \geq 0 : A_m \in B^\beta\}$, the first hitting time of B^β , and $\tilde{\tau}_{B^\alpha} := \inf\{m \geq \tau_{B^\alpha} : A_m \notin B^\alpha\}$, the first leaving time of B^α . Choosing the time horizon $m_0 = \lceil N^{2b+2\varepsilon} \rceil$ with $0 < \varepsilon < 1 - b$, we obtain

$$\begin{aligned} \mathbb{E}[A_{m_0}] &= \mathbb{E}[A_{m_0} | \tilde{\tau}_{B^\alpha} > m_0 \geq \tau_{B^\beta}] \mathbb{P}(\tilde{\tau}_{B^\alpha} > m_0 \geq \tau_{B^\beta}) \\ &\quad + \mathbb{E}[A_{m_0} | m_0 < \tau_{B^\beta}] \mathbb{P}(m_0 < \tau_{B^\beta}) \\ &\quad + \mathbb{E}[A_{m_0} | m_0 > \tilde{\tau}_{B^\alpha} \geq \tau_{B^\beta}] \mathbb{P}(m_0 > \tilde{\tau}_{B^\alpha} \geq \tau_{B^\beta}). \end{aligned}$$

By Lemma 3.7.4 the second summand on the right hand side is of order $O(N \exp(-N^\delta))$ and by Lemma 3.7.3 the third summand is of order $O(N(N^{1-2b+\alpha})^{h_N})$. Concerning the first summand we obtain $\mathbb{P}(\tilde{\tau}_{B^\alpha} > m_0 \geq \tau_{B^\beta}) = 1 - O(N(N^{1-2b+\alpha})^{h_N})$. Observing that $A_{m_0} = \frac{2}{\rho^2}N^{1-b}(1 + o(1))$ whenever $A_{m_0} \in B^\alpha$ we thus conclude

$$\begin{aligned} \mathbb{E}[A_{m_0}] &= \mathbb{E}[A_{m_0} | \tilde{\tau}_{B^\alpha} > m_0 \geq \tau_{B^\beta}] \mathbb{P}(\tilde{\tau}_{B^\alpha} > m_0 \geq \tau_{B^\beta}) + O(N(N^{1-2b+\alpha})^{h_N}) \\ &= \frac{2}{\rho^2}N^{1-b}(1 + o(1)). \end{aligned}$$

This yields the desired upper bound on $\mathbb{E}[A_{\text{eq}}^{(N)}]$. The same argument applies for the lower bound, where we use a CASP started in the state 1 and apply Lemma 3.7.5 instead of Lemma 3.7.4.

2. It remains to show that, as claimed in Theorem 8b), the conditions (3.10) and (3.14) are satisfied in the Wright-Fisher case $W_i^{(N)} \equiv \frac{1}{N}$ (with $\rho^2 = 1$) as well as in the case where $\mathcal{L}(\mathcal{W}^{(N)})$ is a Dirichlet($\alpha_N, \dots, \alpha_N$)-distribution with $\frac{\alpha_N+1}{\alpha_N} = \rho^2 + O(1/N)$ as $N \rightarrow \infty$. For the Wright-Fisher case this is immediate because $\mathbb{E}[(W_1^{(N)})^n] = N^{-n}$. In the symmetric Dirichlet case, $W_1^{(N)}$ is Beta($\alpha_N, (N-1)\alpha_N$)-distributed, with n -th moment $\frac{\Gamma(\alpha_N+n)}{\Gamma(\alpha_N N+n)}$. In particular, the second moment is $\frac{\alpha_N+1}{\alpha_N}$, which because of its assumed convergence gives (3.10). The ratio $\mathbb{E}[(W_1^{(N)})^{n+1}]/\mathbb{E}[(W_1^{(N)})^n]$ equals $\frac{\alpha_N+n}{\alpha_N N+n}$, from which (3.14) is immediate because of the assumed convergence of (α_N) . \square

Chapter 4

Haldane's formula in Cannings models: The case of moderately strong selection

For a class of Cannings models we prove Haldane's formula, $\pi(s_N) \sim \frac{2s_N}{\rho^2}$, for the fixation probability of a single beneficial mutant in the limit of large population size N and in the regime of moderately strong selection, i.e. for $s_N \sim N^{-b}$ and $0 < b < 1/2$. Here, s_N is the selective advantage of an individual carrying the beneficial type, and ρ^2 is the (asymptotic) offspring variance. Our assumptions on the reproduction mechanism allow for a coupling of the beneficial allele's frequency process with slightly supercritical Galton-Watson processes in the early phase of fixation.

4.1 Introduction

Analysing the probability of fixation of a beneficial allele that arises from a single mutant is one of the classical problems in population genetics, see [PW08] for a historical overview. A rule of thumb known as Haldane's formula states that the probability of fixation of a single mutant of beneficial type with small selective advantage $s > 0$ and offspring variance ρ^2 in a large population of individuals, whose total number N is constant over the generations, is approximately equal to $2s/\rho^2$. Originally, this was formulated for the (prototypical) model of Wright and Fisher, in which the next generation arises by a multinomial sampling from the previous one (which leads to $\rho^2 = 1 - \frac{1}{N}$ in the neutral case), with the "reproductive weight" of an individual of beneficial type being increased by the (small) factor $1 + s$. A natural generalization of the Wright-Fisher model are the Cannings models; here one assumes *exchangeable* offspring numbers in the neutral case ([Can74], [Ewe04]), and separately within the sets of all individuals of the beneficial and the non-beneficial type in the selective case ([LL07]).

The reasoning in the pioneering papers by [Fis23], [Hal27] and [Wri31] was based on the insight that, as long as the beneficial type is rare, the number of individuals carrying the beneficial type is a slightly supercritical branching process for which the survival probability is

$$\pi(s) \sim \frac{2s}{\rho^2} \quad \text{as } s \rightarrow 0,$$

where $1 + s$ is the offspring expectation and ρ^2 is the offspring variance (see [Ath92], Theorem 3). The heuristics then is that the branching process approximation should be valid until the beneficial allele has either died out or has reached a fraction of the population that is substantial enough so that the law of large numbers dictates that this fraction should rise to 1.

Notably, [LL07] obtained (for fixed population size N) the result

$$\pi(s) = \frac{1}{N} + \frac{2s}{\rho^2} + o(s) \quad \text{as } s \rightarrow 0, \quad (4.1)$$

as a special case of their explicit analytic representation of $\pi(s)$ within a quite general class of Cannings models and selection mechanisms.

An interesting parameter regime as $N \rightarrow \infty$ is that of *moderate selection*,

$$s_N \sim cN^{-b} \quad \text{with} \quad 0 < b < 1, \quad c > 0, \quad (4.2)$$

which is between the classical regimes of *weak* and *strong* selection. Is the *Haldane asymptotics*

$$\pi(s_N) \sim \frac{2s_N}{\rho^2} \quad \text{as } N \rightarrow \infty, \quad (4.3)$$

valid in the regime (4.2)?

If one could bound in this regime the $o(s)$ -term in (4.1) by $o(N^{-b})$, then (4.1) would turn into (4.3). Such an estimate seems, however, hard to achieve in the analytic framework of [LL07].

The main result of the present paper is a proof of the Haldane asymptotics using an approximation by Galton-Watson processes in the regime of *moderately strong selection*, which corresponds to (4.2) for $0 < b < \frac{1}{2}$. Hereby, we assume that the Cannings dynamics admits a *paintbox representation*, whose random weights are exchangeable and of *Dirichlet-type*, and fulfil a certain moment condition, see Section 4.3. Here, the effect of selection is achieved by a decrease of the reproductive weights of the non-beneficial individuals by the factor $1 - s_N$.

An approximation by Galton-Watson processes was used in [GKWY17] to prove the asymptotics (4.3) in the regime of moderately strong selection for a specific Cannings model that arises in the context of experimental evolution, with the next generation being formed by sampling without replacement from a pool of offspring generated by the parents.

In the case $b \geq \frac{1}{2}$ the method developed in the present paper would fail, because then the Galton-Watson approximation would be controllable only up to a time at which the fluctuations of the beneficial allele (that are caused by the resampling) still dominate the trend that is induced by the selective advantage. However, in [BGPW21a] we proved the Haldane asymptotics (4.3) for the case of *moderately weak selection*, i.e. under Assumption (4.2) with $\frac{1}{2} < b < 1$. There a backward point of view turned out to be helpful, which uses a representation of the fixation probability in terms of sampling duality via the *Cannings ancestral selection graph* developed in [BGPW21a] (see also [GS18]).

The results of the present paper together with those of [BGPW21a] do not cover the boundary case $b = \frac{1}{2}$ between moderately strong and moderately weak selection. We conjecture that the Haldane asymptotics (4.3) is valid also in this case.

4.2 A class of Cannings models with selection

This section is a short recap of [BGPW21a] Section 2; we include it here for self-containedness.

4.2.1 Paintbox representation in the neutral case

Neutral Cannings models are characterized by the exchangeable distribution of the vector $\nu = (\nu_1, \dots, \nu_N)$ of offspring sizes; here the ν_i are non-negative integer-valued random variables which sum to N . An important subclass are the *mixed multinomial* Cannings models. Their offspring size vector ν arises in a two-step manner: first, a vector of random weights $\mathscr{W} = (W_1, \dots, W_N)$ is sampled, which is exchangeable and satisfies $W_1 + \dots + W_N = 1$ and $W_i \geq 0, 1 \leq i \leq N$.

In the second step, a die with N possible outcomes $1, \dots, N$ and outcome probabilities $\mathscr{W} = (W_1, \dots, W_N)$ is thrown N times, and ν_i counts how often the outcome i occurs. Hence, given the random weights \mathscr{W} the offspring numbers $\nu = (\nu_1, \dots, \nu_N)$ are Multinomial(N, \mathscr{W})-distributed. Following Kingman's terminology, we speak of a *paintbox representation* for ν , and call \mathscr{W} the underlying (random) *paintbox*.

This construction is iterated over the generations $g \in \mathbb{Z}$: Let $\mathscr{W}^{(g)} = (W_1^{(g)}, \dots, W_N^{(g)})$ be independent copies of \mathscr{W} , and denote the individuals in generation g by (i, g) , $i \in [N]$. Assume that each individual $(j, g+1)$, $j \in [N] := \{1, \dots, N\}$ in generation $g+1$, chooses its parent in generation g , with conditional distribution

$$\mathbb{P}((i, g) \text{ is the parent of } (j, g+1) | \mathscr{W}^{(g)}) = W_i^{(g)}, \quad \forall i \in [N].$$

where given $\mathscr{W}^{(g)}$ the choices of the parents for individuals $\{(j, g+1), j \in [N]\}$ are independent and identically distributed. This results in exchangeable offspring vectors $\nu^{(g)}$ which are independent and identically distributed over the generations g .

For notational simplicity we do not always display dependence of $\mathscr{W}^{(g)}$ on the generation g , and write \mathscr{W} instead. From time to time however we want to emphasise the dependence of \mathscr{W} on N and therefore write $\mathscr{W}^{(N)}$ instead of \mathscr{W} .

Some exchangeable offspring vectors do not have a paintbox representation, for example a random permutation of the vector $(2, \dots, 2, 0, \dots, 0)$. Prototypical paintboxes are $\mathscr{W} = (\frac{1}{N}, \dots, \frac{1}{N})$, which leads to the Wright-Fisher model, and the class of Dirichlet(α, \dots, α)-distributed random weights. In particular, the offspring distribution with Dirichlet($1, \dots, 1$)-distributed paintbox can be seen as a limiting case of the offspring distribution for the model of experimental evolution considered in ([BGPW19], [GKWY17]).

4.2.2 A paintbox representation with selection

Let $\mathscr{W}^{(g)}$, $g \in \mathbb{Z}$, be as in the previous section, and let $s_N \in [0, 1)$. Assume each individual carries one of two types, either the *beneficial type* or the *wildtype*. Depending on the type of individual (i, g) we set

$$\widetilde{W}_i^{(g)} = (1 - s_N)W_i^{(g)}$$

if (i, g) is of wildtype and $\widetilde{W}_i^{(g)} = W_i^{(g)}$ if (i, g) is of beneficial type. The probability that an individual is chosen as parent is now given by

$$\mathbb{P}((i, g) \text{ is parent of } (j, g + 1)) = \frac{\widetilde{W}_i^{(g)}}{\sum_{\ell=1}^N \widetilde{W}_\ell^{(g)}} \quad (4.4)$$

for all $i, j \in [N]$. Parents are chosen independently for all $i \in [N]$ and the distribution does not change over the generations. If (i, g) is the parent of $(j, g + 1)$ the child $(j, g + 1)$ inherits the type of its parent. In particular, this reproduction mechanism leads to offspring numbers that are exchangeable among the beneficial as well as among wildtype individuals.

4.2.3 The Cannings frequency process

In the previous section we gave a definition for a Cannings model which incorporates selection, by decreasing the random weight of each wildtype individual by the factor $1 - s_N$. This allows to define the Cannings frequency process $\mathcal{X} = (X_g)_{g \geq 0}$ with state space $[N]$ which counts the number of beneficial individuals in each generation g . Assume there are $1 \leq k \leq N$ beneficial individuals at time g ; due to the exchangeability of $\mathscr{W}^{(g)}$ we may assume that the individuals $(1, g), \dots, (k, g)$ are the beneficial ones. Given $\mathscr{W}^{(g)} = \mathscr{W}$, the probability that individual $(j, g + 1)$ is of beneficial type is then due to (4.4) equal to

$$\frac{\sum_{i=1}^k W_i}{\sum_{i=1}^k W_i + (1 - s_N) \sum_{i=k+1}^N W_i}, \quad (4.5)$$

and is the same for all $j \in [N]$. Hence, given $\mathscr{W}^{(g)} = \mathscr{W}$ and given there are k beneficial individuals in generation g , the number of beneficial individuals in generation $g + 1$ has distribution

$$\text{Bin} \left(N, \frac{\sum_{i=1}^k W_i}{\sum_{i=1}^k W_i + (1 - s_N) \sum_{i=k+1}^N W_j} \right); \quad (4.6)$$

this defines the transition probabilities of the Markov chain \mathcal{X} .

4.3 Main Result

Before we state our main result we specify the assumptions on the paintbox and the strength of selection.

Definition 4.3.1. (*Dirichlet-type weights*) We say that a random vector $\mathscr{W}^{(N)}$ with exchangeable components $W_1^{(N)}, \dots, W_N^{(N)}$ is of Dirichlet-type if

$$W_i^{(N)} = \frac{Y_i}{\sum_{\ell=1}^N Y_\ell}, \quad i = 1, \dots, N, \quad (4.7)$$

where Y_1, \dots, Y_N are independent copies of a random variable Y with $\mathbb{P}(Y > 0) = 1$.

We assume that

$$\mathbb{E}[\exp(hY)] < \infty, \quad (4.8)$$

for some $h > 0$, which implies the finiteness of all moments of Y . The relevance (and possible relaxations) of Condition (4.8) are discussed further in Remark 4.3.3 a), see also the comment in Remark 4.3.2 a).

Remark 4.3.2. a) *The biological motivation for considering Dirichlet-type weights comes from seasonal reproductive schemes. At the beginning of a season a set (of size N) of individuals is alive. These individuals and their offspring reproduce and generate a pool of descendants within that season. Only a few individuals from this pool survive till the next season. The number N in the model is assumed to be the total number of individuals that make it to the next season. Dirichlet-type weights arise in the asymptotics of an infinitely large pool of offspring; then sampling with and without replacement coincide. Condition (4.8), which we will require for the proof of Theorem 9 (see also Remark 4.3.3), guarantees that the pool of descendants of a single individual is not too large in comparison to the pool of descendants generated by the other individuals. The simplifying assumption $\mathbb{P}(Y > 0) = 1$ implies that the weight $W_i^{(N)}$ of a parent cannot be equal to zero. Observe, however, that weights of single parents can be arbitrarily small if (e.g.) Y has a density which is continuous and strictly positive in zero.*

b) *The case of a deterministic Y corresponds to $W_i^{(N)} \equiv 1/N$, i.e. the classical Wright-Fisher model. If Y has a Gamma(κ)-distribution, then $\mathscr{W}^{(N)}$ is Dirichlet (κ, \dots, κ)-distributed.*

Theorem 1 in [HM21], gives a classification of a large class of Cannings models with a paintbox of the form (4.7) with regard to the convergence of their rescaled genealogies.

c) *Let $\nu^{(N)}$ be a sequence of Cannings offspring numbers that are represented by the paintboxes $\mathscr{W}^{(N)}$. It is well known (and easily checked) that*

$$\text{Var} \left(\nu_1^{(N)} \right) = N(N-1) \mathbb{E}[(W_1^{(N)})^2]. \quad (4.9)$$

If $\mathscr{W}^{(N)}$ is of the form (4.3.1) with $\mathbb{E}[Y^2] < \infty$, which is clearly implied by (4.8), then (see [HM21] Theorem 1 (i)) the right hand side of (4.9) converges to $\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2}$ as $N \rightarrow \infty$.

In view of Remark 4.3.2 c) we have for the asymptotic neutral offspring variance

$$\lim_{N \rightarrow \infty} \text{Var} \left(\nu_1^{(N)} \right) =: \rho^2 = \frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2}. \quad (4.10)$$

Replacing Y by $Y' = \frac{Y}{\mathbb{E}[Y]}$ does not affect (4.7), hence we can and will assume $\mathbb{E}[Y] = 1$ in the proofs, which simplifies (4.10) to $\rho^2 = \mathbb{E}[Y^2]$ (and makes (4.10) consistent with the notation of [HM21]). Under the assumption $\mathbb{E}[Y^4] < \infty$, which is implied by (4.8) as well, the following asymptotics is valid

$$\text{Var} \left(\nu_1^{(N)} \right) = \rho^2 + O(N^{-1}), \text{ as } N \rightarrow \infty. \quad (4.11)$$

We will discuss the relevance of this asymptotics in Remark 4.3.3 b), and prove it in Lemma 4.4.2 b).

Turning to the selective advantage, we assume that for a fixed $\eta \in (0, \frac{1}{4})$ the sequence (s_N) obeys

$$N^{-\frac{1}{2}+\eta} \leq s_N \leq N^{-\eta}, \quad (4.12)$$

which we call the regime of *moderately strong selection*, thus generalizing the corresponding notion introduced in Section 4.1. (Note that (4.12) has an analogue in the regime of *moderately weak selection* as discussed in [BGPW21a]). In order to connect to (4.2) we define

$$b_N := -\frac{\ln s_N}{\ln N} \quad (4.13)$$

which is equivalent to $s_N = N^{-b_N}$, with (4.12) translating to

$$\eta \leq b_N \leq \frac{1}{2} - \eta.$$

We now state our main result on the asymptotics (as $N \rightarrow \infty$) of the fixation probability of the Cannings frequency process $(X_g^{(N)}) = (X_g)$ defined in Subsection 4.2.3. Note that the Markov chain (X_g) has the two absorbing states 0 and N , with the hitting time of $\{0, N\}$ being a.s. finite for all N .

Theorem 9. (*Haldane's formula*)

Assume that Conditions (4.7), (4.8) and (4.12) are fulfilled. Let $(X_g)_{g \geq 0}$ be the number of beneficial individuals in generation g , with $X_0 = 1$. Let $\tau = \inf \{g \geq 0 : X_g \in \{0, N\}\}$, then

$$\mathbb{P}(X_\tau = N) \sim \frac{2s_N}{\rho^2}, \quad \text{as } N \rightarrow \infty. \quad (4.14)$$

We give the proof of Theorem 9 in Section 4.5, after preparing some auxiliary results in Section 4.4. Next we give a strategy of the proof and its main ideas, with an emphasis on the role of Condition (4.12). In Remark 4.3.3 we discuss possible relaxations of Condition (4.8) and the boundary case $b = \frac{1}{2}$.

The proof of Theorem 9 is divided into three parts, corresponding to three growth phases of \mathcal{X} . Concerning the *first phase* we show that the probability to reach the level $N^{b+\delta}$ is $\frac{2s_N}{\rho^2}(1 + o(1))$, for some small $\delta > 0$ and $b := b_N$; this is the content of Proposition 4.5.3. The proof is based on stochastic domination from above and below by slightly supercritical Galton-Watson processes $\bar{\mathcal{Z}}$ and $\underline{\mathcal{Z}}$ with respective offspring distributions (4.29) and (4.30).

To construct a Galton-Watson stochastic upper bound $\bar{\mathcal{Z}}$ of \mathcal{X} in its initial phase, we recall that the transition probabilities of \mathcal{X} are mixed Binomial specified by (4.6). Using (4.7) we approximate (4.5) from above by

$$\frac{1 + s_N + o(s_N)}{N} \sum_{\ell=1}^k Y_\ell. \quad (4.15)$$

As we will show in Lemma 4.4.4, this is possible with probability $1 - O(\exp(-c' N^{1-2\alpha}))$ for some $\alpha < \frac{1}{2}$ and $c' > 0$, and for $k \leq N^{b+\delta}$ with $b + \delta < 1/2$. We will then be able to dominate the mixed Binomial distribution (4.6) by the mixed Poisson distribution with random parameter (4.15), again up to an error term of order $o(s_N)$.

Noting that (4.15) is a sum of independent random variables, we arrive at the upper Galton-Watson approximation for a single generation. For any small $\varepsilon > 0$ this can be repeated for $N^{b+\varepsilon}$ generations, which (as an application of Lemma 4.4.1 will show) is enough to reach either the level 0 or the level $N^{b+\delta}$ with probability $1 - o(s_N)$.

To obtain a Galton-Watson stochastic lower bound $\underline{\mathcal{Z}}$ of \mathcal{X} in its initial phase, we adapt an approach that was used in [GKWY17] in a related situation. As in Subsection 4.2.1, number the individuals in generation g by (i, g) , now with $(1, g), \dots, (X_g, g)$, being the beneficial individuals, and denote by $\omega_i^{(g)}$ the number of children of the individual (i, g) , $1 \leq i \leq X_g$. As will be explained in the proof of Lemma 4.5.1, as long as X_g has not reached the level $N^{b+\delta}$, the distribution of $\omega_i^{(g)}$ can be bounded from below by a mixed binomial distribution

$$\text{Bin} \left(N - \lceil N^{b+\delta} \rceil, Y_1 \frac{1 + s_N + o(s_N)}{N} \right)$$

with probability $1 - O(\exp(-N^\varepsilon))$ for some sufficiently small $\varepsilon > 0$, again for $b + \delta < 1/2$. A suitable stopping and truncation at the level $N^{b+\delta}$ will give the Galton-Watson process approximation from below for the first phase.

We will verify in Subsection 4.4.1 that both slightly supercritical branching processes $\underline{\mathcal{Z}}$ and $\overline{\mathcal{Z}}$ reach the level $N^{b+\delta}$ with probability $\frac{2s_N}{\rho^2}(1 + o(1))$.

As to the *second phase*, we will argue in Section 4.5.2 that, after reaching the level $N^{b+\delta}$ the Cannings frequency process \mathcal{X} will grow to a macroscopic fraction εN with high probability. If the frequency of beneficial individuals is at least $N^{b+\delta}$ (but still below εN), then in a single generation the frequency of beneficial individuals grows in expectation at least by $1 + (1 - \varepsilon)s_N + o(s_N)$. Hence, $cs_N^{-1} \ln N$ generations after \mathcal{X} has reached the level $N^{b+\delta}$, the expected value of the process \mathcal{X} reaches the level $2\varepsilon N$. Similarly one bounds the variance produced in a single generation and derives from this an estimate for the variance accumulated over $cs_N^{-1} \ln N$ generations. This bound being sufficiently small, an application of Chebyshev's inequality yields that (after $cs_N^{-1} \ln N$ generations) \mathcal{X} crosses the level εN with probability tending to 1 after reaching the level $N^{b+\delta}$.

In Section 4.5.3) we deal with the *last phase*, and will show that the fixation probability tends to 1 as $N \rightarrow \infty$ if we start with at least εN individuals of beneficial type. Here we use the representation for the fixation probability that is based on a sampling duality between the Cannings frequency process and the Cannings ancestral selection process (CASP) which was provided in [BGPW21a]. For a subregime of moderately weak selection the claim will follow quickly from the representation formula combined with a concentration result for the equilibrium distribution of the CASP that was proved in [BGPW21a]. To complete the proof we will then argue that both the CASP and the representation of the fixation probability depend on the selection parameter in a monotone way.

Remark 4.3.3. *a) With some additional work the assumption (4.8) of the existence of some exponential moment of Y can be relaxed to some weaker moment condition. In order not to overload the present paper, we restrict here to a sketch.*

In Lemma 4.5.1 we couple the frequency process of the beneficial individuals with Galton-Watson processes for $N^{b+\delta}$ generations. By means of the estimates in Lemma

4.4.3 and Lemma 4.4.4 we show that these couplings hold for a single generation with probability $1 - O(\exp(-N^{c'}))$ for some appropriate $c' > 0$. Since we need the couplings to hold for $N^{b+\delta}$ generations, it suffices that the couplings hold in a single generation with probability $1 - O(N^{-2(b+\delta)})$ for some $\delta > 0$ (since in this case the probability of the coupling to fail is $o(s_N)$ and therefore can be neglected with regard to (4.14)). Such probability bounds can also be obtained under weaker assumptions on the distribution of the random variable Y . Assume e.g. that Y has a regularly varying tail, i.e. $\mathbb{P}(Y > x) \sim x^{-\beta} L(x)$ for some $\beta > 0$ and L is a slowly varying function. For the proof of Lemma 4.5.1 we need to estimate the probability of the event figuring in Lemma 4.4.3 with $b < c \leq 1$ and the probability of the event figuring in Lemma 4.4.4 with $b < \alpha < \frac{1}{2}$. To show that these probabilities are of order $O(N^{-2(b+\delta)})$ we only need that $\mathbb{P}(\sum_{i=1}^n Y_i > x) = O(n^{-2(b+\delta)})$ (since the remaining probability in Lemma 4.4.3 can be estimated with Hoeffding's inequality, see [Hoe94]) with $n = N, x = N^{1-\alpha}$ in Lemma 4.4.4 and $n = N^c, x = N^c$ in Lemma 4.4.3. The asymptotics (3.2) in [MN98] states that $\mathbb{P}(\sum_{i=1}^n Y_i > x) \sim nx^{-\beta} L(x)$. Consequently, we need to choose $\beta > 0$ such that $N^{1-\beta(1-\alpha)} L(N^{1-\alpha}) = O(N^{-2(b+\delta)})$ as well as $N^{c-\beta c} L(N^c) = O(N^{-2(b+\delta)})$. This works for all choices of $0 < b < \frac{1}{2}$, provided that $\beta \geq 4$.

It would be nice to have a proof of the asymptotics (4.14) under the assumption that the 4th moment of Y is finite, even without the assumption of a regularly varying tail.

The investigation of the analogue to (4.14) in the absence of finite second moments, i.e. for Cannings models with heavy-tailed offspring distributions, is the subject of ongoing research, and will be treated in a forthcoming paper.

- b) Relation (4.11) will be used in the proof of Lemma 4.5.4. Moreover, this relation is also instrumental in the companion paper [BGPW21a] (on the regime of moderately weak selection). The special case $n = 3$ in Lemma 4.4.2 a) shows that the assumption $\mathbb{E}[Y^3] < \infty$ implies

$$\mathbb{E}[(W_1^{(N)})^3] = O(N^{-3}).$$

This gives a rate of decay $O(N^{-2})$ for the triple coalescence probability (and is the moment condition (3.6) in [BGPW21a]).

Condition (4.8) (on the existence of an exponential moment of Y) guarantees the Haldane asymptotics (4.14) for Cannings models with weights of Dirichlet type also in the whole regime of moderately weak selection $N^{-1+\eta} \leq s_N \leq N^{-\frac{1}{2}-\eta}$ without any further assumption. In particular the assumption on the finiteness of a negative moment of Y in [BGPW21a], Lemma 3.7 b), is unnecessary. Indeed, in the proof of Lemma 4.4.2 a) we show that $\mathbb{E}[(W_1^{(N)})^n] \leq \left(\frac{2}{N}\right)^n (\mathbb{E}[Y^n] + o(1))$. As shown in the proof of Lemma 3.7 b) in [BGPW21a] Condition (4.8) guarantees that for a sequence (h_N) with $h_N \rightarrow \infty$ and $h_N \in O(\log N)$ for all $n \leq 2h_N$ we can estimate $\mathbb{E}[Y^n]$ from above by $C \left(\frac{2h_N}{c}\right)^n$ for appropriate constants $C, c > 0$. Consequently, for N sufficiently large we have $\mathbb{E}[(W_1^{(N)})^n] \leq \left(\frac{Kh_N}{N}\right)^n$ for some appropriate constant $K > 0$, that is Condition (3.8) in [BGPW21a] is fulfilled.

- c) It seems a mathematically intriguing question whether in the regime of moderate selection all Cannings models which admit a paintbox representation with Dirichlet-type weights and are in Kingman domain of attraction, also follow the Haldane asymptotics (4.14).

An example of a sequence of Cannings models (with weights not of Dirichlet-type) which fulfil Möhle's condition but do not follow the Haldane asymptotics, is the following. In each generation a randomly chosen individual gets weight $N^{-\gamma}$, $0 < \gamma < \frac{1}{2}$ and all the other individuals have a weight of $\frac{1-N^{-\gamma}}{N-1}$. Then we have $\mathbb{E}[W_1^2] \sim N^{-1-2\gamma}$ and $\mathbb{E}[W_1^3] = o(\mathbb{E}[W_1^2])$, therefore by Möhle's criterion the genealogy lies in the attraction of Kingman's coalescent. However, the Haldane asymptotics would predict that the survival probability is of order $s_N/(N^2 N^{-1-2\gamma}) \sim N^{-1-b+2\gamma}$, which for $\gamma < b/2$ is $\ll N^{-1}$. Since the fixation probability of a beneficial allele cannot be smaller than the fixation probability under neutrality (which is $\frac{1}{N}$), (4.14) must be violated in this example.

- d) The present work together with the approach in [BGPW21a] does not cover the boundary case $b = \frac{1}{2}$. A quick argument why our arguments cannot be extended simply to the boundary case is the following. We show that once the beneficial type exceeds (in the order of magnitude) the frequency $s_N^{-1} = N^b$ it goes to fixation with high probability. In the regime $b < \frac{1}{2}$ we use couplings with Galton-Watson processes to show that this threshold is reached with probability $\frac{2s_N}{\rho^2}(1 + o(1))$. However, these couplings are not guaranteed as soon as collisions occur, i.e. when beneficial individuals are replacing beneficial individuals. By the well known "birthday problem" collisions are common as soon as $N^{\frac{1}{2}}$ individuals are of the beneficial type. Therefore we require $N^b \ll N^{\frac{1}{2}}$, i.e. $b < \frac{1}{2}$.

In the light of the results of the present paper and of [BGPW21a], there is little reason to doubt that the assertion of Theorem 9 should fail in the boundary case $b = 1/2$. However, the question remains open (and intriguing) whether then the backward or the forward approach (or a combination of both) is the appropriate tool for the proof.

4.4 Auxiliary results

4.4.1 Slightly supercritical Galton-Watson processes

Throughout this subsection, $(s_N)_{N \in \mathbb{N}}$ is a sequence of positive numbers converging to 0, σ^2 is a fixed positive number, and $Z^{(N)} = (Z_n^{(N)})_{n \geq 0}$, $N = 1, 2, \dots$ are Galton-Watson processes with offspring expectation

$$\mathbb{E}_1[Z_1^{(N)}] = 1 + s_N + o(s_N), \quad (4.16)$$

offspring variance $\sigma^2 + o(1)$ and uniformly bounded third moments $\mathbb{E}_1[(Z_1^{(N)})^3]$. Unless stated otherwise we assume that $Z_0^{(N)} = 1$. We write

$$\phi_N := \mathbb{P}\left(\lim_{n \rightarrow \infty} Z_n^{(N)} = \infty\right) = 1 - \mathbb{P}\left(Z_n^{(N)} = 0 \text{ for some } n > 1\right) \quad (4.17)$$

for the survival probability of $(Z^{(N)})$ and observe

$$\phi_N \sim \frac{2s_N}{\sigma^2}. \quad (4.18)$$

The derivation and discussion of the asymptotics (4.18) has a venerable history, a few key references being [Hal27], [Kol38], [Esh81], [Hop92], [Ath92] Theorem 3, [HJV05] Theorem 5.5.

Lemma B.3 in [GKWW17] gives a statement on the asymptotic probability that $Z^{(N)}$ either quickly dies out or reaches a certain (moderately) large threshold. The

following lemma improves on this in a twofold way. It dispenses with the assumption $s_N \sim cN^{-b}$ for a fixed $b \in (0, 1)$ and more substantially, it gives a *quantitative* estimate for the probability that, given non-extinction, the (moderately) large threshold is reached quickly.

Lemma 4.4.1. *Fix $\delta > 0$, and let $T^{(N)} := \inf\{n \geq 0 : Z_n^{(N)} \notin \{1, 2, \dots, \lceil (\frac{1}{s_N})^{1+\delta} \rceil\}$. Then, for all $\varepsilon > 0$*

$$\mathbb{P}_1 \left(T^{(N)} > (1/s_N)^{1+\varepsilon} \right) = O \left(\exp \left(-cs_N^{-\varepsilon/2} \right) \right), \quad (4.19)$$

with $c = -\log \left(\frac{7}{8} \right)$.

Proof. Observe that

$$\begin{aligned} \mathbb{P}_1 \left(T^{(N)} > (1/s_N)^{1+\varepsilon} \right) &\leq \mathbb{P}_1 \left(T^{(N)} > (1/s_N)^{1+\varepsilon} \mid Z^N \text{ survives} \right) \\ &\quad + \mathbb{P}_1 \left(T^{(N)} > (1/s_N)^{1+\varepsilon} \mid Z^N \text{ dies out} \right). \end{aligned} \quad (4.20)$$

In Part 1 of the proof we will estimate the first probability on the r.h.s. of (4.20); this will give the above-mentioned improvement of Lemma B.3 in [GKWY17]. Part 2 of the proof deals with the second probability on the r.h.s. of (4.20).

Part 1. Like in the proof of Lemma B.3 in [GKWY17] we obtain an upper bound on the time at which the process $Z^{(N)}$ reaches the level $(1/s_N)^{1+\delta}$ given survival, by considering the process $Z^* = (Z_n^*)_{n \geq 0}$ consisting of the immortal lines of $Z^{(N)}$ conditioned to non-extinction. (For simplicity of notation we drop a superscript N in Z^* .) Let ϕ_N denote the survival probability of $Z^{(N)}$ as in (4.17). The offspring distribution of Z^* arises from that of $Z^{(N)}$ as

$$\mathbb{P}_1(Z_1^* = k) = \frac{1}{\phi_N} \mathbb{E}_1 \left[\binom{Z_1^{(N)}}{k} \phi_N^k (1 - \phi_N)^{Z_1^{(N)} - k} \right], \quad k \geq 1. \quad (4.21)$$

(see [LP17] Proposition 5.28). In particular one has

$$\mathbb{E}_1[Z_1^*] = \frac{1}{\phi_N} \mathbb{E}_1[Z_1^{(N)} \phi_N] = \mathbb{E}_1[Z_1^{(N)}].$$

Denote, as usual, for a random variable X and an event A by $\mathbb{E}[X; A] := \mathbb{E}[X \mathbb{1}_A]$. Furthermore,

$$\mathbb{E}_1[Z_1^*; Z_1^* \geq 3] \leq \frac{1}{\phi_N} \mathbb{E}_1 \left[\binom{Z_1^{(N)}}{3} \phi_N^3 \right] = O(\phi_N^2)$$

because of the assumed uniform boundedness of the third moments of $Z_1^{(N)}$. These two relations together with (4.16), (4.18) and the fact that $\mathbb{E}_1[Z_1^*; Z_1^* = 1] \leq 1$ immediately give a lower bound for $\mathbb{E}_1[Z_1^*; Z_1^* = 2] \leq 1$, implying that for any $\beta \in (0, 1)$

$$\mathbb{P}_1(Z_1^* \geq 2) \geq \beta s_N, \quad \mathbb{P}_1(Z_1^* = 1) \leq 1 - \beta s_N. \quad (4.22)$$

Hence the process $Z^{(N)}$, when conditioned on survival, is bounded from below by the counting process Z^* of immortal lines, which in turn is bounded from below by the process $\tilde{Z} = (\tilde{Z}_n)_{n \geq 0}$ with offspring distribution

$$\nu = (1 - \beta s_N) \delta_1 + \beta s_N \delta_2.$$

So far we closely followed the proof in [GKWY17], but now we deviate from that proof to obtain the rate of convergence claimed in (4.19).

An upper bound for the time $\tilde{T} := \inf\{n \geq 0 : \tilde{Z}_n \geq (1/s_N)^{1+\delta}\}$ also gives an upper bound for the time $T^{(N)}$. The idea is now to divide an initial piece of $k \leq (1/s_N)^{(1+\varepsilon)}$ generations into $\lfloor (1/s_N)^{\varepsilon/2} \rfloor$ parts, each of $n_0 \leq (1/s_N)^{(1+\varepsilon/2)}$ generations. Because of the immortality of \tilde{Z} and the independence between these parts we obtain immediately that

$$\mathbb{P}(\tilde{T} \geq (1/s_N)^{(1+\varepsilon)}) \leq \mathbb{P}_1(\tilde{Z}_j \leq (1/s_N)^{1+\delta} \text{ for } j = 1, \dots, k) \leq \left(\mathbb{P}_1(\tilde{Z}_{n_0} \leq (1/s_N)^{1+\delta}) \right)^{\lfloor (1/s_N)^{\varepsilon/2} \rfloor}$$

We then bound $\mathbb{P}_1(\tilde{Z}_{n_0} > (1/s_N)^{1+\delta})$ from below by an application of the Paley-Zygmund inequality in its form

$$\mathbb{P}\left(X \geq \frac{\mathbb{E}[X]}{2}\right) \geq \frac{1}{4} \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}, \quad (4.23)$$

where X is a non-negative random variable (with finite second moment). For a supercritical Galton-Watson process with offspring expectation m and offspring variance σ^2 the n -th generation offspring expectation and n -th generation offspring variance σ_n^2 are given by m^n and $\sigma^2 m^n (m^n - 1)/(m^2 - m)$ (see [AN72], p.4). Hence, we obtain

$$\mathbb{E}_1[\tilde{Z}_n] = (1 + \beta s_N)^n, \quad \text{Var}_1[\tilde{Z}_n] = \frac{\beta s_N (1 - \beta s_N) (1 + \beta s_N)^n ((1 + \beta s_N)^n - 1)}{(1 + \beta s_N)^2 - (1 + \beta s_N)} \quad (4.24)$$

We choose the smallest n_0 such that

$$\mathbb{E}_1[\tilde{Z}_{n_0}] \geq 2(1/s_N)^{1+\delta}. \quad (4.25)$$

Observe that $n_0 \sim \frac{1}{\beta s_N} \log(2(1/s_N)^{1+\delta})$ which ensures that $(1/s_N)^{\varepsilon/2} n_0 \leq (1/s_N)^{1+\varepsilon}$ for N large enough. We now estimate $\mathbb{E}_1[(\tilde{Z}_{n_0})^2]$ using (4.24) as follows

$$\begin{aligned} \mathbb{E}\left[\tilde{Z}_{n_0}^2\right] &= \frac{\beta s_N (1 - \beta s_N) (1 + \beta s_N)^{n_0} ((1 + \beta s_N)^{n_0} - 1)}{(1 + \beta s_N)^2 - (1 + \beta s_N)} + (1 + \beta s_N)^{2n_0} \\ &\leq \frac{\beta s_N (1 + \beta s_N)^{2n_0}}{\beta s_N + (\beta s_N)^2} + (1 + \beta s_N)^{2n_0} \leq 2(1 + \beta s_N)^{2n_0}. \end{aligned}$$

Applying (4.23) with $X := \tilde{Z}_{n_0}$ yields

$$\mathbb{P}_1\left(\tilde{Z}_{n_0} \geq \frac{1}{2} \mathbb{E}\left[\tilde{Z}_{n_0}\right]\right) \geq \frac{1}{4} \frac{(1 + \beta s_N)^{2n_0}}{2(1 + \beta s_N)^{2n_0}} = \frac{1}{8},$$

which because of (4.25) implies $\mathbb{P}_1(\tilde{Z}_{n_0} \leq (1/s_N)^{1+\delta}) \leq \frac{7}{8}$. If after time n_0 the process \tilde{Z}_{n_0} is still smaller than our desired bound $(1/s_N)^{1+\delta}$, we can iterate this argument $\lfloor (1/s_N)^{\varepsilon/2} \rfloor$ times and arrive at

$$\left(\mathbb{P}_1(\tilde{Z}_{n_0} \leq (1/s_N)^{1+\delta})\right)^{\lfloor (1/s_N)^{\varepsilon/2} \rfloor} \leq \left(\frac{7}{8}\right)^{\lfloor (1/s_N)^{\varepsilon/2} \rfloor} = \exp(-c \lfloor (1/s_N)^{\varepsilon/2} \rfloor),$$

with $c = -\log \frac{7}{8}$. This gives the desired bound for the first term in (4.20).

Part 2. We now turn to the second term on the r.h.s. of (4.20). Define

$$T_0^{(N)} := \inf\{n \geq 0 : Z_n^{(N)} = 0\}.$$

Obviously $T^{(N)} \leq T_0^{(N)}$, and so it suffices to prove

$$\mathbb{P}\left(T_0^{(N)} > (1/s_N)^{1+\varepsilon} \mid Z^{(N)} \text{ dies out}\right) \leq \exp(-\beta s_N^{-\varepsilon}(1+o(1))). \quad (4.26)$$

This proof follows closely that of the second part of Lemma B.3 in [GKWY17]; we include it here for completeness.

We observe

$$\begin{aligned} \mathbb{E}_1 \left[Z_1^{(N)} \mid Z^{(N)} \text{ dies out} \right] &= \frac{1}{1-\phi_N} \mathbb{E}_1 \left[(1-\phi_N)^{Z_1^{(N)}} Z_1^{(N)} \right] = \mathbb{E}_1 \left[(1-\phi_N)^{Z_1^{(N)}-1} Z_1^{(N)} \right] \\ &= \mathbb{P}_1(Z_1^* = 1), \end{aligned}$$

where the first and the last equality follow from the branching property and from (4.21), respectively. We have shown in (4.22) that $\mathbb{P}_1(Z_1^* = 1) \leq 1 - \beta s_N + o(s_N)$ and hence we can conclude

$$\mathbb{E} \left[Z_{\lfloor (1/s_N)^{1+\varepsilon} \rfloor} \mid Z^{(N)} \text{ dies out} \right] \leq (1 - \beta s_N + o(s_N))^{\lfloor (1/s_N)^{1+\varepsilon} \rfloor} \leq \exp(-\beta s_N^{-\varepsilon}(1+o(1))).$$

Finally, an application of Markov's inequality yields (4.26)

$$\mathbb{P}(T_0^{(N)} > (1/s_N)^{1+\varepsilon} \mid Z^{(N)} \text{ dies out}) \leq \mathbb{P}(Z_{\lfloor (1/s_N)^{1+\varepsilon} \rfloor} \geq 1 \mid Z^{(N)} \text{ dies out}) \leq \exp(-\beta s_N^{-\varepsilon}(1+o(1))).$$

□

4.4.2 Estimates on the paintbox

The following lemma provides the asymptotics (4.11) as well as the moment bounds for the Dirichlet-type weights that were addressed in Remark 4.3.3 b).

Lemma 4.4.2 (Moments of the weights). *Let Y, Y_1, \dots, Y_N be iid positive random variables with $\mathbb{E}[Y] = 1$ and $\rho^2 = \mathbb{E}[Y^2]$. We abbreviate $H_N := Y_1 + \dots + Y_N$.*

a) *Assume $\mathbb{E}[Y^n] < \infty$ for some $n \in \mathbb{N}$. Then*

$$\mathbb{E} \left[\left(\frac{Y}{(Y + H_N)} \right)^n \right] = O(N^{-n}).$$

b) *Assume $\mathbb{E}[Y^4] < \infty$. Then*

$$\mathbb{E} \left[\left(\frac{Y}{(Y + H_N)} \right)^2 \right] = \frac{\rho^2}{N^2} + O(N^{-3}).$$

Proof. a) Consider the event $F_N := \{\frac{H_N}{N} \leq \frac{1}{2}\}$. First we note that

$$\mathbb{E} \left[\left(\frac{Y}{Y + H_N} \right)^n \right] \leq \mathbb{E} \left[\left(\frac{Y}{Y + H_N} \right)^n \mathbb{1}_{F_N} \right] + \mathbb{E} \left[\left(\frac{Y}{H_N} \right)^n \mathbb{1}_{F_N^c} \right] \leq \mathbb{P}(F_N) + \left(\frac{2}{N} \right)^n \mathbb{E}[Y^n].$$

Let K be so large that $\mathbb{E}[Y \wedge K] > \frac{1}{2}$. Hoeffding's inequality applied to the sample mean of i.i.d. copies of the bounded random variable $Y \wedge K$ implies that $\mathbb{P}(F_N)$ decays exponentially fast. Since $\mathbb{E}[Y^n]$ is bounded this yields the claim.

b) We observe that

$$\mathbb{E} \left[\frac{Y^2}{(Y + H_N)^2} \right] = \frac{1}{N^2} \mathbb{E} \left[\frac{Y^2}{\left(\frac{Y}{N} + \frac{H_N}{N} \right)^2} \right].$$

Let $E_N^{(1)} := \{\frac{H_N}{N} > \frac{5}{4}\}$, $E_N^{(2)} := \{\frac{H_N}{N} < \frac{3}{4}\}$ and $E_N^{(3)} := \{Y > \frac{N}{4}\}$. Markov's inequality applied to $\mathbb{P} \left(\left(\sum_{i=1}^N (Y_i - 1) \right)^4 \geq \frac{N^4}{4^4} \right)$ together with the assumption $\mathbb{E}[Y^4] < \infty$ implies $\mathbb{P}(E_N^{(1)}) = O(N^{-2})$. Likewise, $\mathbb{P}(E_N^{(3)}) = O(N^{-4})$. Furthermore, let K be so large that $\mathbb{E}[Y \wedge K] > \frac{3}{4}$. Again, Hoeffding's inequality applied to the sample mean of i.i.d. copies of the bounded random variable $Y \wedge K$ together with monotonicity imply that $\mathbb{P}(E_N^{(2)})$ decays exponentially; a fortiori we have $\mathbb{P}(E_N^{(2)}) = O(N^{-3})$.

Let $E_N := E_N^{(1)} \cup E_N^{(2)} \cup E_N^{(3)}$. We have $\mathbb{E} \left[\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N} \right)^2} \mathbb{1}_{E_N} \right] = O(N^{-1})$ since on $E_N^{(1)}$ we have $\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N} \right)^2} \leq Y^2$ and on $E_N^{(2)}$ and $E_N^{(3)}$ we have $\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N} \right)^2} \leq N^2$. Hence, it remains to show that $\mathbb{E} \left[\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N} \right)^2} \mathbb{1}_{E_N^c} \right] = \rho^2 + O(N^{-1})$. Define $Z_N := \frac{1}{\sqrt{N}}(H_N - N)$ and observe

$$\mathbb{E} \left[\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N} \right)^2} \mathbb{1}_{E_N^c} \right] = \mathbb{E} \left[\frac{Y^2}{\left(1 + \frac{Z_N}{\sqrt{N}} + \frac{Y}{N} \right)^2} \mathbb{1}_{E_N^c} \right].$$

Abbreviate $R_N = 2 \left(\frac{Z_N}{\sqrt{N}} + \frac{Y}{N} \right) + \left(\frac{Z_N}{\sqrt{N}} + \frac{Y}{N} \right)^2$. On E_N^c we have $-\frac{1}{2} \leq R_N$ and hence

$$1 - R_N \leq \frac{1}{\left(\frac{H_N}{N} + \frac{Y}{N} \right)^2} = \frac{1}{1 + R_N} \leq 1 - R_N + 2R_N^2.$$

Thus

$$\mathbb{E}[Y^2(1 - R_N)\mathbb{1}_{E_N^c}] \leq \mathbb{E} \left[\frac{Y^2}{\left(\frac{H_N - N}{N} + \frac{Y}{N} \right)^2} \mathbb{1}_{E_N^c} \right] \leq \mathbb{E}[Y^2(1 - R_N + 2R_N^2)\mathbb{1}_{E_N^c}] \quad (4.27)$$

By Cauchy-Schwarz we have $\mathbb{E}[Y^2\mathbb{1}_{E_N^c}] = \mathbb{E}[Y^2] + O(N^{-1})$. Similarly, $\mathbb{E}[Y^2 Z_N \mathbb{1}_{E_N}] = \mathbb{E}[Y^2 Z_N] + O(N^{-1}) = O(N^{-1})$, since $\mathbb{E}[Y^2 Z_N]$ vanishes due to the independence of Y and Z_N . The remaining terms in (4.27) are $O(N^{-1})$ as well, which completes the proof of Lemma 4.4.2. \square

We now prove a bound on the deviations for the total weight of k individuals.

Lemma 4.4.3. (*Large deviations bound for a moderate number of random weights*)

Let (Y_i) and $(W_i^{(N)})$ satisfy (4.7), (4.8), $\mathbb{E}[Y] = 1$ and let $k = k_N \leq N^c$ for some $0 < c \leq 1$. Then for all $\varepsilon > 0$ there exists a positive constant c_ε depending only on ε and the distribution of Y such that

$$\mathbb{P} \left(\sum_{i=1}^k W_i^{(N)} \geq (1 + \varepsilon)N^{c-1} \right) = O(\exp(-c_\varepsilon N^c)). \quad (4.28)$$

Proof. This follows by a combination of two Cramér bounds. Indeed, the l.h.s. of (4.28) is by assumption bounded from above by

$$\mathbb{P} \left(\frac{\sum_{i=1}^{\lceil N^c \rceil} Y_i}{\sum_{j=1}^N Y_j} \geq (1 + \varepsilon)N^{c-1} \right).$$

Abbreviating $E := \{\sum_{j=1}^N Y_j \geq (1 - \varepsilon')N\}$ with ε' such that $(1 + \varepsilon)(1 - \varepsilon') > 1$ we estimate the latter probability from above by

$$\begin{aligned} & \mathbb{P} \left(\sum_{i=1}^{\lceil N^c \rceil} Y_i \geq (1 + \varepsilon)N^{c-1} \sum_{j=1}^N Y_j, E \right) + \mathbb{P}(E^c) \\ & \leq \mathbb{P} \left(\sum_{i=1}^{\lceil N^c \rceil} Y_i \geq N^c(1 + \varepsilon)(1 - \varepsilon') \right) + \mathbb{P}(E^c) \\ & = O(e^{-N^c I((1+\varepsilon)(1-\varepsilon'))}) + O(e^{-NI(1-\varepsilon')}), \end{aligned}$$

denoting by $I(y)$ the rate function of Y . Due to (4.8) $I(y)$ exists around $\mathbb{E}[Y] = 1$ and is strictly positive for $y \neq 1$ (see [DZ94] Theorem 2.2.3). This yields an upper bound of $O(\exp(-c_\varepsilon N^c))$ with $c_\varepsilon = \min\{I((1 + \varepsilon)(1 - \varepsilon')), I(1 - \varepsilon')\}$. \square

The next lemma gives stochastic upper and lower bounds for the sums of the random weights in terms of sums of the independent random variables Y_i .

Lemma 4.4.4. (*Bounds for the random weights*)

Assume that Conditions (4.7) and (4.8) are fulfilled and $\mathbb{E}[Y] = 1$. Let $0 < \alpha < \frac{1}{2}$, then for $k = k_N \leq N$

$$\mathbb{P} \left(\frac{1 - N^{-\alpha}}{N} \sum_{i=1}^k Y_i \leq \sum_{i=1}^k W_i^{(N)} \leq \frac{1 + N^{-\alpha}}{N} \sum_{i=1}^k Y_i \right) \geq 1 - \exp(-c' N^{1-2\alpha}) (1 + o(1)),$$

for some $c' > 0$.

Proof. It suffices to show

$$\mathbb{P} \left(\left| \frac{N}{\sum_{j=1}^N Y_j} - 1 \right| \geq N^{-\alpha} \right) = O(\exp(-N^{1-2\alpha})).$$

For $0 < c < 1$ we have

$$\mathbb{P}\left(\sum_{i=1}^N Y_i < cN\right) = O(\exp(-NI(c))),$$

where $I(y)$ is the rate function of Y . Condition (4.8) ensures that $I(c) > 0$ for $\mathbb{E}[Y_1] = 1 \neq c$, see [DZ94] Theorem 2.2.3.

For any $a, a' \geq 1$ one has $|\frac{1}{a} - \frac{1}{a'}| \leq |a - a'|$. This yields

$$\begin{aligned} \mathbb{P}\left(\left|\frac{N}{\sum_{j=1}^N Y_j} - 1\right| \geq N^{-\alpha}\right) &= \mathbb{P}\left(\frac{1}{c} \left|\frac{Nc}{\sum_{j=1}^N Y_j} - c\right| \geq N^{-\alpha}\right) \\ &\leq \mathbb{P}\left(\frac{1}{c^2} \left|\frac{\sum_{j=1}^N Y_j}{N} - 1\right| \geq N^{-\alpha}\right) + O(e^{-NI(c)}) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{N}} \left|\sum_{i=1}^N (Y_i - 1)\right| \geq c^2 N^{\frac{1}{2}-\alpha}\right) + O(e^{-NI(c)}). \end{aligned}$$

Using [Cra38] Theorem 1, the probability on the r.h.s. can, with a suitable $\tilde{c} > 0$, be estimated from above by

$$\exp(\tilde{c}N^{1-3\alpha}) \exp\left(-\frac{c^4}{2}N^{1-2\alpha}\right)(1 + O(N^{-\alpha} \log N)) = \exp\left(-\frac{c^4}{2}N^{1-2\alpha}\right)(1 + o(1)),$$

which gives the desired result. \square

4.5 Proof of the main result

Recall from (4.13) that we denote the order of the selection strength by $b_N = -\frac{\log s_N}{\log N}$. To simplify notation we will drop the subscript and simply write $b := b_N$. As mentioned already in the sketch of the proof of Theorem 9 we assume without loss of generality that $\mathbb{E}[Y] = 1$.

The proof of the Theorem is divided into three parts, which correspond to three phases of growth for the Cannings frequency process \mathcal{X} . The initial phase is decisive: due to Proposition 4.5.3, the probability that \mathcal{X} reaches the level $N^{b+\delta}$ for some sufficiently small δ is given by the r.h.s. of (4.14). Lemma 4.5.4 and Lemma 4.5.5 then guarantee that, once having reached the level $N^{b+\delta}$, the process \mathcal{X} reaches N with high probability. The proof of the Theorem is then a simple combination of these three results and the strong Markov property. Indeed, with τ_1, τ_2, τ_3 as in Proposition 4.5.3, Lemma 4.5.4 and Lemma 4.5.5, and with $\delta, \delta', \varepsilon$ fulfilling the requirements specified there, the fixation probability in the l.h.s. of (4.14) can be rewritten as

$$\begin{aligned} \mathbb{P}(X_\tau = N) &= \mathbb{P}(X_{\tau_3} = N | X_{\tau_2} \geq \varepsilon N) \mathbb{P}(X_{\tau_2} \geq \varepsilon N | X_{\tau_1} \geq N^{b+\delta}) \mathbb{P}_1(X_{\tau_1} \geq N^{b+\delta}) \\ &= (1 - o(1))(1 - O(N^{-\delta'})) \frac{2s_N}{\rho^2} (1 + o(1)) \\ &\sim \frac{2s_N}{\rho^2}. \end{aligned}$$

4.5.1 First phase: From 1 to $N^{b+\delta}$

In this section we show that as long as $X_g \leq N^{b+\delta}$ the process \mathcal{X} can be upper and lower bounded (with sufficiently high probability) by two slightly supercritical branching processes $\underline{\mathcal{Z}} = (\underline{Z}_g)_{g \geq 0}$ and $\overline{\mathcal{Z}} = (\overline{Z}_g)_{g \geq 0}$. To construct the upper bound $\overline{\mathcal{Z}}$ we take the highest per capita selective advantage, which occurs when only a single individual is beneficial. Using Lemma 4.4.3 and Lemma 4.4.4, we will approximate the thus arising mixed binomial distribution by a mixed Poisson distribution, which leads for $\overline{\mathcal{Z}}$ to the offspring distribution

$$\text{Pois}(Y_1(1 + s_N + o(s_N))), \quad (4.29)$$

where Y_1 is the random variable figuring in (4.7). To arrive at the lower bounding Galton-Watson process $\underline{\mathcal{Z}}$ we note that the per capita selective advantage is bounded from below by the one when $\lceil N^{b+\delta} \rceil$ beneficial individuals are present in the parent generation, as long as the process \mathcal{X} has not reached the level $N^{b+\delta}$. Again using Lemma 4.4.3 and Lemma 4.4.4 we will show that the offspring distribution of $\underline{\mathcal{Z}}$ can be chosen as the mixed binomial distribution

$$\text{Bin}\left(N - \lceil N^{b+\delta} \rceil, \frac{Y_1}{N}(1 + s_N + o(s_N))\right). \quad (4.30)$$

Lemma 4.5.1. (*Coupling with Galton-Watson processes*)

Let δ and α be such that $0 < \delta < \eta$ and $\frac{1}{2} - \eta < \alpha < \frac{1}{2}$, and put

$$\tau_1 = \inf\{g \geq 0 : X_g = 0 \text{ or } X_g \geq N^{b+\delta}\}.$$

Then \mathcal{X} can be defined on one and the same probability space together with two branching process $\underline{\mathcal{Z}}$ and $\overline{\mathcal{Z}}$ with offspring distributions (4.30) and (4.29), respectively, such that for $j = 1, 2, \dots$

$$\begin{aligned} \mathbb{P}(\underline{Z}_{j \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil \leq X_{j \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil \leq \overline{Z}_{j \wedge \tau_1} \mid \underline{Z}_{j-1 \wedge \tau_1} \leq X_{j-1 \wedge \tau_1} \leq \overline{Z}_{j-1 \wedge \tau_1}) \\ \geq 1 - e^{-c'N^{1-2\alpha}}(1 + o(1)), \end{aligned} \quad (4.31)$$

with c' as in Lemma 4.4.4.

Applying the latter estimate g times consecutively yields immediately the following corollary:

Corollary 4.5.2. Let $\delta, \alpha, \tau_1, \underline{\mathcal{Z}}$ and $\overline{\mathcal{Z}}$ be as in Lemma 4.5.1. If $X_0 \leq N^{b+\delta}$, then for all $g \in \mathbb{N}_0$

$$\mathbb{P}(\underline{Z}_{g \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil \leq X_{g \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil \leq \overline{Z}_{g \wedge \tau_1} \mid \underline{Z}_0 \leq X_0 \leq \overline{Z}_0) \geq (1 - O(\exp(-c'N^{1-2\alpha})))^g. \quad (4.32)$$

Proof. of Lemma 4.5.1. We proceed inductively, assuming that for $g = 1, 2, \dots$ we have constructed \mathcal{X} , $\overline{\mathcal{Z}}$ and $\underline{\mathcal{Z}}$ up to generation $g-1$ such that (4.31) holds for $j = 1, \dots, g-1$. Together with X_g we will construct \overline{Z}_g and \underline{Z}_g , and check the asserted probability bound for the coupling.

Given $\{X_{g-1} = k\}$ and the weights (W_i) in generation $g-1$, the number of beneficial individuals X_g in generation g has the binomial distribution (4.6). Aiming first at the construction of the upper bound $\overline{\mathcal{Z}}$, we relate (4.6) to (4.29) in terms of stochastic order. For $p, p' \geq 0$, a $\text{Bin}(N, p)$ -distributed random variable B is stochastically

dominated by a $\text{Pois}(Np')$ -distributed random variable P if

$$e^{-p'} \leq (1 - p), \quad (4.33)$$

see (1.21) in [KM10]. Indeed, in this case the probability of the outcome zero is not larger for a $\text{Pois}(p')$ -distributed random variable P_1 than for a $\text{Bernoulli}(p)$ -distributed random variable B_1 , which yields $B_1 \preceq P_1$, where \preceq denotes the usual stochastic ordering of the random variables. Consequently

$$B \stackrel{d}{=} \sum_{i=1}^N B_i \preceq \sum_{i=1}^N P_i \stackrel{d}{=} P.$$

with B_i and P_i being independent copies of B_1 and P_1 , respectively. In particular, for $p \geq 0$ and $p' = p(1 + N^{b+2\delta-1})$ we have

$$e^{-p'} \leq 1 - p' + (p')^2 = 1 - p(1 + N^{b+2\delta-1}) + p^2(1 + N^{b+2\delta-1})^2.$$

Hence Condition (4.33) holds if

$$p(1 + N^{b+2\delta-1})^2 < N^{b+2\delta-1}. \quad (4.34)$$

Given (W_i) , the success probability of the binomial distribution (4.6) is bounded from above via

$$p := \left(\sum_{i=1}^k W_i \right) / (1 - s_N).$$

Thus by Lemma 4.4.3, (4.34) is fulfilled with probability $1 - O(\exp(-c_\varepsilon N^{b+\delta}))$ with c_ε as in Lemma 4.4.3. In this sense the number of beneficial offspring is dominated by a $\text{Pois}\left(N \frac{\sum_{i=1}^k W_i}{(1 - s_N)} (1 + N^{b+2\delta-1})\right)$ -distributed random variable with high probability. Applying Lemma 4.4.4 yields that with probability $1 - \exp(-c' N^{1-2\alpha})(1 + o(1))$ the following chain of inequalities is valid:

$$\begin{aligned} N \frac{\sum_{i=1}^{X_{g-1}} W_i}{(1 - s_N)} (1 + N^{b+2\delta-1}) &\leq \frac{\sum_{i=1}^{X_{g-1}} Y_i}{(1 - s_N)} (1 + N^{b+2\delta-1}) (1 + N^{-\alpha}) \\ &= \sum_{i=1}^{X_{g-1}} Y_i (1 + s_N + o(s_N)) \leq \sum_{i=1}^{\bar{Z}_{g-1}} Y_i (1 + s_N + o(s_N)). \end{aligned}$$

In this way \mathcal{X} can be coupled with a branching process $\bar{\mathcal{Z}}$ with a mixed Poisson offspring distribution of the form (4.29).

The lower bound also uses a comparison with a Galton-Watson process, now with a mixed binomially distributed offspring distribution:

Number the individuals in generation $g-1$ by $(i, g-1)$, with $(1, g-1), \dots, (X_{g-1}, g-1)$, being the beneficial individuals. Given \mathscr{W} , we use a sequence of coin tossings to determine which of the individuals from generation g are the children of $(i, g-1)$. The first N tosses determine which individuals are the children of $(1, g-1)$. Denoting the number of these children by ω_1^{g-1} , the next $N - \omega_1^{g-1}$ tosses (with an updated success probability) determine which individuals are the children of $(2, g-1)$, etc. Observe that as long as $X_{g-1} \leq N^{b+\delta}$, and given \mathscr{W} and $\sum_{\ell=1}^{i-1} \omega_\ell^{(g-1)} =: h$, then $\omega_i^{(g-1)}$

for $i \leq X_{g-1}$ has distribution

$$\text{Bin} \left(N - h, \frac{W_i}{\sum_{\ell=i}^{X_{g-1}} W_\ell + (1 - s_N) \sum_{\ell=X_{g-1}+1}^N W_\ell} \right). \quad (4.35)$$

Note that the success probability in (4.35) can be estimated from below by

$$\frac{W_i}{\sum_{\ell=1}^{X_{g-1}} W_\ell + (1 - s_N) \sum_{\ell=X_{g-1}+1}^N W_\ell} = \frac{W_i}{1 - s_N + s_N \sum_{\ell=1}^{X_{g-1}} W_\ell}.$$

As long as $X_{g-1} \leq \lceil N^{b+\delta} \rceil$, Lemma 4.4.3 ensures that for $\varepsilon > 0$

$$\frac{W_i}{1 - s_N + s_N \sum_{j=1}^{X_{g-1}} W_j} \geq \frac{W_i}{1 - s_N + (1 + \varepsilon)N^{\delta-1}} \quad (4.36)$$

with probability $1 - O(\exp(-c_\varepsilon N^{b+\delta}))$. Lemma 4.4.4, in turn, yields that the r.h.s. of (4.36) is bounded from below by

$$\frac{Y_i}{N(1 - s_N + (1 + \varepsilon)N^{\delta-1})} (1 + N^{-\alpha}) = \frac{Y_i}{N} (1 + s_N + o(s_N))$$

with probability at least $1 - \exp(-c' N^{1-2\alpha})(1 + o(1))$.

Thus, if $\omega_1^{(g-1)} + \dots + \omega_{i-1}^{(g-1)} = h \leq \lceil N^{b+\delta} \rceil$, then the distribution of $\omega_i^{(g-1)}$ specified in (4.35) is bounded from below by

$$\text{Bin} \left(N - \lceil N^{b+\delta} \rceil, \frac{W_i}{1 - s_N + (1 + \varepsilon)N^{\delta-1}} \right)$$

with probability $1 - O(\exp(-c_\varepsilon N^{b+\delta}))$.

If $\omega_1^{(g-1)} + \dots + \omega_{i-1}^{(g-1)} = h > \lceil N^{b+\delta} \rceil$, then we have $\underline{Z}_{g \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil \leq X_{g \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil$. Consequently \mathcal{X} can be coupled with a Galton-Watson process \underline{Z} with offspring distribution of the form (4.30) such that also the lower estimate in (4.32) is fulfilled. This completes the proof of Lemma 4.5.1. \square

We are now ready to prove that \mathcal{X} reaches the level $N^{b+\delta}$ with probability $\frac{2s_N}{\rho^2}(1 + o(1))$.

Proposition 4.5.3. (*Probability to reach the critical level*)

Assume Conditions (4.7), (4.8) and (4.12) are fulfilled and define $\tau_1 = \inf\{g \geq 0 : X_g \geq N^{b+\delta} \text{ or } X_g = 0\}$ with $0 < \delta < \eta$, then

$$\mathbb{P}(X_{\tau_1} \geq N^{b+\delta}) = \frac{2s_N}{\rho^2}(1 + o(1)). \quad (4.37)$$

Proof. We use the couplings of \mathcal{X} with the slightly supercritical branching processes \underline{Z} and \bar{Z} from Corollary 4.5.2 and show that both processes reach the level $N^{b+\delta}$ with probability $\frac{2s_N}{\rho^2}(1 + o(1))$. Let $\delta' > 0$ and E be the event that the stochastic ordering between \underline{Z} , \mathcal{X} and \bar{Z} holds until generation $n_0 = \lceil N^{b+\delta'} \rceil$, that is

$$E = \{\underline{Z}_0 \wedge \lceil N^{b+\delta'} \rceil \leq X_0 \wedge \lceil N^{b+\delta'} \rceil \leq \bar{Z}_0, \dots, \underline{Z}_{n_0} \wedge \lceil N^{b+\delta'} \rceil \leq X_{n_0} \wedge \lceil N^{b+\delta'} \rceil \leq \bar{Z}_{n_0}\}.$$

We show below that the stopping time τ_1 fulfils

$$\mathbb{P}(\tau_1 \geq \lceil N^{b+\delta'} \rceil) = o(s_N). \quad (4.38)$$

For some g that is polynomially bounded in N , the r.h.s. of (4.32) is bounded from above by $1 - o(s_N)$. Thus, combining Corollary 4.5.2 and (4.38) we deduce

$$\mathbb{P}(E, \tau_1 \leq \lceil N^{b+\delta'} \rceil) = 1 - o(s_N).$$

We are now going to bound (4.37) from above by estimating the corresponding probability for $\bar{\mathcal{Z}}$ and the stopping time $\bar{\tau}_1 = \inf\{g \geq 0 : \bar{Z}_g \geq N^{b+\delta} \text{ or } \bar{Z}_g = 0\}$. More precisely,

$$\begin{aligned} \mathbb{P}(X_{\tau_1} \geq N^{b+\delta}) &= \mathbb{P}(X_{\tau_1} \geq N^{b+\delta}, \tau_1 \leq \lceil N^{b+\delta'} \rceil, E) + o(s_N) \\ &\leq \mathbb{P}(\bar{Z}_{\bar{\tau}_1} \geq N^{b+\delta}, \tau_1 \leq \lceil N^{b+\delta'} \rceil, E) + o(s_N) \\ &\leq \mathbb{P}(\bar{Z}_{\bar{\tau}_1} \geq N^{b+\delta}) + o(s_N). \end{aligned}$$

To obtain an upper bound for the probability of $\bar{\mathcal{Z}}$ to reach the level $N^{b+\delta}$ it suffices to estimate the survival probability of $\bar{\mathcal{Z}}$. For notational simplicity let us write $\{\bar{\mathcal{Z}} \text{ survives}\}$ for the event $\{\forall g \geq 0 : \bar{Z}_g > 0\}$ and similarly $\{\bar{\mathcal{Z}} \text{ dies out}\}$ for the event $\{\exists g \geq 0 : \bar{Z}_g = 0\}$. We have

$$\begin{aligned} &\mathbb{P}_1(\bar{Z}_{\bar{\tau}_1} \geq N^{b+\delta}) \\ &\leq \mathbb{P}_1(\bar{Z}_{\bar{\tau}_1} \geq N^{b+\delta} | \bar{\mathcal{Z}} \text{ survives}) \mathbb{P}_1(\bar{\mathcal{Z}} \text{ survives}) + \mathbb{P}_1(\bar{Z}_{\bar{\tau}_1} \geq N^{b+\delta} | \bar{\mathcal{Z}} \text{ dies out}) \\ &\leq \mathbb{P}_1(\bar{\mathcal{Z}} \text{ survives}) + \mathbb{P}(\text{all } \lceil N^{b+\delta} \rceil \text{ individuals die out}) = \mathbb{P}_1(\bar{\mathcal{Z}} \text{ survives}) + (1 - \mathbb{P}_1(\bar{\mathcal{Z}} \text{ survives}))^{\lceil N^{b+\delta} \rceil}. \end{aligned}$$

The survival probability of $\bar{\mathcal{Z}}$ will now be estimated by means of (4.18). To this purpose we calculate the expectation and variance of the offspring distribution (4.29). The expectation is $1 + s_N + o(s_N)$ and the variance is given by

$$\begin{aligned} &\text{Var}(\text{Pois}(Y_1(1 + s_N + o(s_N)))) \\ &= \text{Var}\left(\mathbb{E}\left[\text{Pois}(Y_1(1 + s_N + o(s_N))) \mid Y_1\right]\right) + \mathbb{E}\left[\text{Var}\left(\text{Pois}(Y_1(1 + s_N + o(s_N))) \mid Y_1\right)\right] \\ &= \text{Var}(Y_1(1 + s_N + o(s_N))) + \mathbb{E}[Y_1(1 + s_N + o(s_N))] \\ &= (1 + s_N + o(s_N))^2 \text{Var}(Y_1) + 1 + s_N + o(s_N) = \rho^2(1 + o(1)). \end{aligned}$$

Equation (4.18) yields that the survival probability of the process $\bar{\mathcal{Z}}$ is given by $\frac{2s_N}{\rho^2}(1 + o(1))$. The lower bound in (4.37) follows by similar arguments by considering the process $\underline{\mathcal{Z}}$ instead.

It remains to show (4.38). Define $\bar{\tau}^{(0)}$ and $\underline{\tau}^{(u)}$ as the stopping times that the process $\bar{\mathcal{Z}}$ reaches 0 and the process $\underline{\mathcal{Z}}$ reaches the upper bound $N^{b+\delta}$, respectively, i.e.

$$\bar{\tau}^{(0)} = \inf\{g \geq 0 : \bar{Z}_g = 0\}, \quad \underline{\tau}^{(u)} = \inf\{g \geq 0 : \underline{Z}_g \geq N^{b+\delta}\},$$

with the convention that the infimum over an empty set is infinity. Then

$$\begin{aligned}
\mathbb{P}(\tau_1 \geq \lceil N^{b+\delta'} \rceil) &\leq \mathbb{P}(\tau_1 \geq \lceil N^{b+\delta'} \rceil, E) + \mathbb{P}(E^c) \\
&\leq \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'} \rceil, \bar{\tau}^{(0)} \geq \lceil N^{b+\delta'} \rceil, E) + \mathbb{P}(E^c) \\
&= \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'} \rceil, \bar{\tau}^{(0)} \geq \lceil N^{b+\delta'} \rceil, E, \underline{\mathcal{Z}} \text{ dies out}) \\
&\quad + \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'} \rceil, \bar{\tau}^{(0)} \geq \lceil N^{b+\delta'} \rceil, E, \underline{\mathcal{Z}} \text{ survives}) + \mathbb{P}(E^c) \\
&= \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'} \rceil, \bar{\tau}^{(0)} \geq \lceil N^{b+\delta'} \rceil, E, \underline{\mathcal{Z}} \text{ dies out}) \\
&\quad + O(e^{-c'N^{\frac{\delta'}{2}}}) + O(N^{b+\delta'} e^{-\frac{1}{2}N^{1-2\alpha}}), \tag{4.39}
\end{aligned}$$

by an application of Lemma 4.4.1 and Corollary 4.5.2 and $\alpha < \frac{1}{2}$ as defined there. To keep the notation simple we denote by e_N terms of the order $\exp(-N^c)$ for some $c > 0$. Proceeding with (4.39) we obtain

$$\begin{aligned}
(4.39) &= \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'} \rceil, \bar{\tau}^{(0)} \geq \lceil N^{b+\delta'} \rceil, E, \underline{\mathcal{Z}} \text{ dies out}, \bar{\mathcal{Z}} \text{ survives}) \\
&\quad + \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'} \rceil, \bar{\tau}^{(0)} \geq \lceil N^{b+\delta'} \rceil, E, \underline{\mathcal{Z}} \text{ dies out}, \bar{\mathcal{Z}} \text{ dies out}) + e_N \\
&\leq \mathbb{P}(\underline{\mathcal{Z}} \text{ dies out}, \bar{\mathcal{Z}} \text{ survives}, E) + e_N, \tag{4.40}
\end{aligned}$$

again by an application of Lemma 4.4.1. Note that

$$\begin{aligned}
&\mathbb{P}(\underline{\mathcal{Z}} \text{ dies out}, \bar{\mathcal{Z}} \text{ survives}, E) + \mathbb{P}(\underline{\mathcal{Z}} \text{ survives}, \bar{\mathcal{Z}} \text{ survives}, E) \\
&= \mathbb{P}(\bar{\mathcal{Z}} \text{ survives}, E) = \frac{2s_N}{\rho^2}(1 + o(1)).
\end{aligned}$$

In order to show that (4.40) is $o(s_N)$ it suffices to prove that

$$\mathbb{P}(\underline{\mathcal{Z}} \text{ survives}, \bar{\mathcal{Z}} \text{ survives}, E) = \frac{2s_N}{\rho^2}(1 + o(1)).$$

Considering again the event $\{\underline{\tau}^{(u)} \leq \lceil N^{b+\delta'} \rceil\}$ and applying (4.18) one obtains

$$\begin{aligned}
\mathbb{P}(\underline{\mathcal{Z}} \text{ survives}, \bar{\mathcal{Z}} \text{ survives}, E) &= \mathbb{P}(\underline{\mathcal{Z}} \text{ survives}, \bar{\mathcal{Z}} \text{ survives}, E, \underline{\tau}^{(u)} \leq \lceil N^{b+\delta'} \rceil) + e_N \\
&= \mathbb{P}(\underline{\mathcal{Z}} \text{ survives}, E, \underline{\tau}^{(u)} \leq \lceil N^{b+\delta'} \rceil) + e_N,
\end{aligned}$$

since the events E and $\{\underline{\tau}^{(u)} \leq \lceil N^{b+\delta'} \rceil\}$ imply that $\bar{Z}_g \geq N^{b+\delta}$ for some $g \leq N^{b+\delta'}$ and the probability for $\bar{\mathcal{Z}}$ to die out after reaching $N^{b+\delta}$ is $(1 - \frac{2s_N}{\rho^2}(1 + o(1)))^{N^{b+\delta}} = e_N$. One more application of (4.18) yields

$$\mathbb{P}(\underline{\mathcal{Z}} \text{ survives}, E, \underline{\tau}^{(u)} \leq \lceil N^{b+\delta'} \rceil) = \frac{2s_N}{\rho^2}(1 + o(1)),$$

which finishes the proof. \square

4.5.2 Second phase: from $N^{b+\delta}$ to εN

In this section we show that \mathcal{X} , once having reached the level $N^{b+\delta}$, will reach the level εN with probability tending to 1 as $N \rightarrow \infty$.

Lemma 4.5.4. (From $N^{b+\delta}$ to εN with high probability)

Assume $X_0 \geq N^{b+\delta}$ with $0 < \delta < \eta$, let $0 < \varepsilon < \frac{\delta}{2-2\eta-\delta}$ and define the stopping time

$$\tau_2 = \inf\{g \geq 0 : X_g \notin \{1, 2, \dots, \lfloor \varepsilon N \rfloor\}\}.$$

Then there exists some $\delta' > 0$ such that

$$\mathbb{P}(X_{\tau_2} \geq \varepsilon N) = 1 - O(N^{-\delta'}).$$

Proof. By monotonicity it is enough to prove the claim for $X_0 = \lceil N^{b+\delta} \rceil$. By definition we have

$$\mathcal{L}(X_{g+1}|X_g) = \text{Bin} \left(N, \frac{\sum_{i=1}^{X_g} W_i}{\sum_{i=1}^{X_g} W_i + (1-s_N) \sum_{i=X_g+1}^N W_i} \right). \quad (4.41)$$

Next we lower-bound \mathcal{X} by the process $\tilde{\mathcal{X}} = (\tilde{X}_g)_{g \geq 0}$, $\tilde{X}_0 = X_0$, with conditional distribution

$$\mathcal{L}(\tilde{X}_{g+1}|\tilde{X}_g) = \text{Bin} \left(N, \frac{\sum_{i=1}^{\tilde{X}_g} W_i}{1-s_N \sum_{i=\varepsilon N+1}^N W_i} \right) \quad (4.42)$$

as long as $\tilde{X}_g \leq \varepsilon N$. If $\tilde{X}_g > \varepsilon N$ we assume that \tilde{X}_{g+1} is distributed as a slightly supercritical branching process with $\text{Pois}(Y_{1q_N})$ distributed offspring, where

$$q_N = N \mathbb{E} \left[\frac{W_1}{1-s_N \sum_{i=\varepsilon N+1}^N W_i} \right]. \quad (4.43)$$

We will see that by this definition in each generation the expectation of $\tilde{\mathcal{X}}$ increases by the factor q_N , see (4.45) and (4.46). The generation-wise increase of the variance conditioned on the current state can be estimated from above by a factor $\rho^2(1+o(1))$, see (4.45) and (4.47), leading to an iterative estimate on the variance of the form (4.48). As long as $X_g \geq \tilde{X}_g$, the success probability in the mixed Binomial distribution on the r.h.s. of (4.41) dominates the corresponding one on the r.h.s. of (4.42). Consequently, starting $\tilde{\mathcal{X}}$ and X both in $\lfloor N^{b+\delta} \rfloor$ we can couple them, such that $\tilde{X}_g \leq X_g$ as long as $\tilde{\mathcal{X}}$ did not cross the level εN . In particular, we have for $\tilde{\tau} = \inf\{g \geq 0 : \tilde{X}_g \notin \{1, 2, \dots, \varepsilon N\}\}$

$$\mathbb{P}(X_{\tau} \geq \varepsilon N) \geq \mathbb{P}(\tilde{X}_{\tilde{\tau}} \geq \varepsilon N). \quad (4.44)$$

To show $\mathbb{P}(\tilde{X}_{\tilde{\tau}} \geq \varepsilon N) = 1 - O(N^{-\delta'})$ we will estimate the first and second moment of \tilde{X}_{g_0} for a suitably chosen $g_0 \in \mathbb{N}$ and then use Chebyshev's inequality to show that \tilde{X}_{g_0} is above εN with sufficiently high probability. For this purpose we consider $m(x)$ and $v(x)$, the one-step conditional expectation and variance of \tilde{X} at $x \in \mathbb{N}$, that is

$$m(x) = \mathbb{E} \left[\tilde{X}_1 | \tilde{X}_0 = x \right], \quad v(x) = \text{Var} \left(\tilde{X}_1 | \tilde{X}_0 = x \right).$$

From the definition of $\tilde{\mathcal{X}}$ as a branching process above εN we have for $x > \varepsilon N$

$$m(x) = q_N x, \quad v(x) = \rho^2 x(1 + o(1)). \quad (4.45)$$

Next we show that $m(x)$ and $v(x)$ fulfil relations similar to (4.45) also for $x \leq \varepsilon N$, which will allow to estimate the expectation and the variance of \tilde{X}_{g_0} .

For $x \leq \varepsilon N$ we have due to (4.42)

$$\begin{aligned} m(x) &= N \mathbb{E} \left[\frac{\sum_{i=1}^x W_i}{1 - s_N \sum_{i=\varepsilon N+1}^N W_i} \right] \\ &= N \mathbb{E} \left[\sum_{i=1}^x W_i \left(1 + s_N \sum_{i=\varepsilon N+1}^N W_i + O(s_N^2) \right) \right] \\ &= x(1 + s_N(1 - \varepsilon)N^2 \mathbb{E}[W_1 W_{\varepsilon N+1}] + O(s_N^2)) \\ &= x(1 + s_N(1 - \varepsilon)(1 + O(N^{-1})) + O(s_N^2)) \\ &= x(1 + (1 - \varepsilon)s_N + O(s_N^2)). \end{aligned}$$

In the penultimate equality we used $\mathbb{E}[W_1 W_2] = \frac{1}{N^2} + O(N^{-3})$ which results from the fact

$$1 = \mathbb{E} \left[\left(\sum_{i=1}^N W_i \right)^2 \right] = N \mathbb{E}[W_1^2] + N(N-1) \mathbb{E}[W_1 W_2]$$

and (4.11). Consequently, we have for all $x \in \mathbb{N}$, recalling (4.43),

$$m(x) = x q_N = x(1 + (1 - \varepsilon)s_N + O(s_N^2)). \quad (4.46)$$

Next we analyse $v(x)$, again for $x \leq \varepsilon N$. In view of (4.42), a decomposition of the variance gives

$$\begin{aligned} v(x) &= \text{Var} \left(N \frac{\sum_{i=1}^x W_i}{1 - s_N \sum_{i=\varepsilon N+1}^N W_i} \right) + \mathbb{E} \left[N \frac{\sum_{i=1}^x W_i}{1 - s_N \sum_{i=\varepsilon N+1}^N W_i} \left(1 - \frac{\sum_{i=1}^x W_i}{1 - s_N \sum_{i=\varepsilon N+1}^N W_i} \right) \right]^2 \\ &\leq \mathbb{E} \left[\left(N \frac{\sum_{i=1}^x W_i}{1 - s_N \sum_{i=\varepsilon N+1}^N W_i} \right)^2 \right] - \mathbb{E} \left[N \frac{\sum_{i=1}^x W_i}{1 - s_N \sum_{i=\varepsilon N+1}^N W_i} \right]^2 + \mathbb{E} \left[N \frac{\sum_{i=1}^x W_i}{1 - s_N} \right]^2 \\ &\leq \mathbb{E} \left[\left(N \frac{\sum_{i=1}^x W_i}{1 - s_N} \right)^2 \right] - \mathbb{E} \left[N \sum_{i=1}^x W_i \right]^2 + \mathbb{E} \left[N \frac{\sum_{i=1}^x W_i}{1 - s_N} \right]^2. \end{aligned}$$

Because of the negative correlation of the W_i , the sum of the first and the second term is not larger than $xN^2 \text{Var}(W_1)$, which because of (4.11) is $\leq x(\rho^2 - 1) + O(N^{-1})$. Since the third term is $x(1 + O(s_N))$, we have for all $x \leq \varepsilon N$

$$v(x) \leq \rho^2 x(1 + o(1)). \quad (4.47)$$

Combining (4.46) and (4.47) allows us to estimate the variance $\text{Var}(\tilde{X}_g)$ for $g \in \mathbb{N}$, again by decomposing the variance:

$$\begin{aligned} \text{Var}(\tilde{X}_g) &= \text{Var}\left(\mathbb{E}\left[\tilde{X}_g|\tilde{X}_{g-1}\right]\right) + \mathbb{E}\left[\text{Var}\left(\tilde{X}_g|\tilde{X}_{g-1}\right)\right] \\ &= \text{Var}\left(m(\tilde{X}_{g-1})\right) + \mathbb{E}\left[v(\tilde{X}_{g-1})\right] \\ &\leq q_N^2 \text{Var}\left(\tilde{X}_{g-1}\right) + \rho^2 \mathbb{E}\left[\tilde{X}_{g-1}\right] (1 + o(1)) \\ &= q_N^2 \text{Var}\left(\tilde{X}_{g-1}\right) + \rho^2 q_N^{g-1} \tilde{X}_0 (1 + o(1)). \end{aligned} \quad (4.48)$$

Iterating this argument yields

$$\begin{aligned} \text{Var}(\tilde{X}_g) &= \rho^2 \tilde{X}_0 q_N^{g-1} \sum_{j=0}^{g-1} q_N^j (1 + o(1)) \\ &= \rho^2 \tilde{X}_0 q_N^{g-1} \frac{q_N^g - 1}{q_N - 1} (1 + o(1)). \end{aligned}$$

Choose the minimal $g_0 \in \mathbb{N}$ such that $2\varepsilon N \leq \mathbb{E}\left[\tilde{X}_{g_0}\right] = q_N^{g_0} X_0$, which yields recalling the initial condition $X_0 = \lceil N^{b+\delta} \rceil$

$$g_0 = \left\lceil \frac{\log(2\varepsilon N X_0^{-1})}{\log q_N} \right\rceil = \left\lceil \frac{\log(2\varepsilon N^{1-b-\delta})}{(1-\varepsilon)s_N + O(s_N^2)} \right\rceil.$$

Applying Chebyshev's inequality with $\tilde{X}_0 = X_0$, we obtain

$$\begin{aligned} \mathbb{P}\left(|\tilde{X}_{g_0} - \mathbb{E}\left[\tilde{X}_{g_0}\right]| \geq \varepsilon N\right) &\leq \frac{\rho^2 \tilde{X}_0 q_N^{g_0-1} \frac{q_N^{g_0}-1}{q_N-1} (1 + o(1))}{\varepsilon^2 N^2} \\ &\leq \frac{\rho^2 N^{b+\delta} q_N^{2g_0} \frac{N^b}{(1-\varepsilon)} (1 + o(1))}{\varepsilon^2 N^2} \\ &= \frac{\rho^2}{\varepsilon^2 (1-\varepsilon)} N^{2b+\delta-2} (1 + (1-\varepsilon)s_N + O(s_N^2))^{2g_0} (1 + o(1)) \\ &\leq c_{\rho,\varepsilon} N^{2b+\delta-2} \exp(2g_0 s_N (1 + O(s_N))) (1 + o(1)) \\ &\leq c_{\rho,\varepsilon} N^{2b+\delta-2} N^{\frac{2}{1-\varepsilon}(1-b-\delta)} (1 + o(1)) \\ &= O(N^{-\delta'}), \end{aligned}$$

for some small $\delta' > 0$ due to the assumptions on ε .

Since $\mathbb{E}\left[\tilde{X}_{g_0}\right] \geq 2\varepsilon N$, this implies

$$\mathbb{P}\left(\tilde{X}_{g_0} \geq \varepsilon N\right) \geq \mathbb{P}\left(\tilde{X}_{g_0} \geq \varepsilon N\right) \geq 1 - O(N^{-\delta'})$$

and due to (4.44) this finishes the proof. \square

4.5.3 Third phase: from εN to N

Lemma 4.5.5 below concerns the last step of the proof, showing that once the process \mathcal{X} has reached the level $\lfloor \varepsilon N \rfloor$, it goes to fixation with high probability. Our proof relies on a representation of the fixation probability of \mathcal{X} in terms of (a functional of) the equilibrium state $A_{\text{eq}} := A_{\text{eq}}^{(N)}$ of the counting process $\mathcal{A} := \mathcal{A}^{(N)} = (A_m)_{m \geq 0}$ of the *potential ancestors* in the time discrete Cannings ancestral selection graph as provided by [BGPW21a]. The process $\mathcal{A}^{(N)}$ is called *Cannings ancestral selection process (CASP)* in [BGPW21a]; for fixed N , it is a recurrent, $[N]$ -valued Markov chain whose transition probabilities are specified in [BGPW21a] Section 2.3.

Theorem 3.1 and Formula (3.2) (see also Corollary 3.3) in [BGPW21a] provide the following sampling duality representation of the fixation probability of \mathcal{X} when started with k individuals:

$$\mathbb{P}_k(\mathcal{X} \text{ eventually hits } N) = 1 - \mathbb{E} \left[\frac{(N-k)(N-k-1) \cdots (N-k-A_{\text{eq}}+1)}{N(N-1) \cdots (N-A_{\text{eq}}+1)} \right] \quad (4.49)$$

Intuitively, this says that \mathcal{X} goes extinct if and only if a random sample of (random) size A_{eq} , drawn without replacement from the population of size N , avoids the k beneficial individuals.

Formula (4.49) implies

$$\mathbb{P}_{\lfloor \varepsilon N \rfloor}(\mathcal{X} \text{ eventually hits } N) \geq 1 - \mathbb{E} \left[(1 - \varepsilon)^{A_{\text{eq}}^{(N)}} \right]. \quad (4.50)$$

The representation of the transition probabilities of \mathcal{A} in [BGPW21a] Section 2.3 in terms of two half steps yields that for fixed N CASPs with different selection parameters can be coupled in such a way that $A_{\text{eq}}^{(N)}$ is increasing in s_N . Take a sequence (\tilde{s}_N) satisfying $\tilde{s}_N \leq s_N$ and Condition (1.2) in [BGPW21a], i.e.

$$N^{-1+\eta} \leq \tilde{s}_N \leq N^{-2/3+\eta}.$$

Let $\tilde{A}_{\text{eq}}^{(N)}$ be the equilibrium state belonging to \tilde{s}_N (and to the same Dirichlet-type paintbox as that of \mathcal{X}). The central limit result [BGPW21a], Corollary 6.10, implies that $\tilde{A}_{\text{eq}}^{(N)} \rightarrow \infty$ in probability as $N \rightarrow \infty$. Because of the just mentioned monotonicity in the selection coefficient, the same convergence holds true for the sequence $(A_{\text{eq}}^{(N)})$. The following lemma is thus immediate from (4.50) and dominated convergence:

Lemma 4.5.5. *(From εN to N with high probability)*

Let \mathcal{X} be a Cannings frequency process with $X_0 = k \geq \varepsilon N$ for some $0 < \varepsilon < 1/2$. Assume that Conditions (4.7), (4.8) and (4.12) are fulfilled. Define $\tau_3 := \inf\{g \geq 0 : X_g \in \{0, N\}\}$. Then

$$\mathbb{P}_k(X_{\tau_3} = N) = 1 - o(1).$$

4.6 Discussion

The analysis of fixation probabilities of slightly beneficial mutants is at the heart of population genetics; some seminal and more modern references are given in the Introduction. Our main result concerns Haldane's asymptotics (4.3) for the fixation probability in a regime of moderate selection.

Our framework is that of Cannings models with selection (as reviewed in Section 4.2), where the corresponding neutral genealogies are assumed to belong to the domain of attraction of Kingman's coalescent. This class of models is motivated by seasonal reproduction cycles in which within each season a large number of offspring is generated but only a comparatively small number (concentrated around a carrying capacity N) of randomly sampled offspring survive to the next season. In this setting it is reasonable to approximate sampling without replacement by sampling with replacement. Thus, under the assumption of neutrality, the probability that the j -th offspring that survives till the next generation is a child of parent i is approximately given by the random weight

$$W_i = \frac{Y_i}{\sum_{\ell=1}^N Y_\ell},$$

where Y_1, \dots, Y_N are the sizes of (potential one-generation) offspring of parents $1, \dots, N$. These sizes are assumed to be independent and identically distributed in the present paper, leading to the concept of weights of *Dirichlet type*. The subsequent generation then arises by a multinomial sampling with random weights, and to add selection the weights of wildtype parents are decreased by the factor $(1 - s_N)$. For a closely related model with a specific distribution of Y_i (and sampling without replacement) in the context of Lenski's long-term evolution experiment see [GKWY17] and [BGPW19].

We prove Haldane's asymptotics in the case of *moderately strong selection*, see Theorem 9, in which the selection strength s_N obeys

$$N^{-\frac{1}{2}+\eta} \leq s_N \leq N^{-\eta}$$

for some $\eta > 0$ and a large population size N . In the companion paper [BGPW21a] the range of *moderately weak selection* was considered, i.e. in the case

$$N^{-1+\eta} \leq s_N \leq N^{-\frac{1}{2}-\eta}$$

for some $\eta > 0$. Since $s_N \gg N^{-1}$ in the regime of moderate selection, selection acts in this case on a faster timescale than genetic drift.

In [BGPW21a] an ancestral selection graph for the just described class of Cannings models with selection was defined, and it was shown that the fixation probability π_N is equal to the expected value $\mathbb{E} \left[\frac{A_{\text{eq}}^{(N)}}{N} \right]$, where $A_{\text{eq}}^{(N)}$ is the number of lines of the ancestral selection graph in its equilibrium. While we could analyse directly the asymptotics of that quantity in the regime of moderately weak selection, we were facing too large fluctuations of $A_{\text{eq}}^{(N)}$ in the regime of moderately strong selection in order to be successful with this approach. Conversely, it turned out that the classical idea of branching process approximation is suitable precisely in that latter regime.

For highly skewed offspring distributions an asymptotics for the fixation probability arises which is different from (4.3). In cases where the neutral genealogy is attracted by a Beta($2 - \alpha, \alpha$)-coalescent, [OH21] argue that the fixation probability is proportionally to $s_N^{\frac{1}{\alpha-1}}$, if $1 \gg s_N \gg N^{-(\alpha-1)}$. Thus the probability of fixation is substantially smaller than in Haldane's asymptotics, which is reasonable since the offspring variance is diverging as $N \rightarrow \infty$. Notably, since the evolutionary timescale of Cannings models in the domain of attraction of Beta-coalescents is of the order $N^{\alpha-1}$, the case $1 \gg s_N \gg N^{-(\alpha-1)}$ corresponds to the regime of moderate selection;

note also that the case of coalescents being in the domain of attraction of a Kingman coalescent corresponds formally to $\alpha = 2$.

Acknowledgements

A.G.C. acknowledges CONACyT Ciencia Basica A1-S-14615 and travel support through the SPP 1590 and C.P. and A.W. acknowledge partial support through DFG project WA 967/4-2 in the SPP 1590. We thank Götz Kersting for fruitful discussions, and we are grateful to two anonymous reviewers for their very helpful comments.

Conflict of interest

The authors declare that they have no conflict of interest.

Chapter 5

Haldane's asymptotics for Supercritical Branching Processes in an iid Random Environment

Branching processes in a random environment are a natural generalisation of Galton-Watson processes. In this paper we analyse the asymptotic decay of the survival probability for a sequence of slightly supercritical branching processes in a random environment, where the offspring expectation converges from above to 1. We prove that Haldane's asymptotics, known from classical Galton-Watson processes, turns up again in the random environment case, provided that one stays away from the critical or subcritical regimes. A central building block is a connection to and a limit theorem for perpetuities with asymptotically vanishing interest rates.

Keywords and Phrases. branching process, random environment, supercriticality, perpetuity, survival probability

MSC 2010 subject classification. Primary 60J80, Secondary 92D25

5.1 Introduction and main result

In the early twentieth century Fisher [Fis23], Haldane [Hal27] and Wright [Wri31] studied the survival probability of a beneficial mutant gene in large populations. They argued that as long as the mutant is sufficiently rare and the selective advantage is small the number of mutants should evolve like a slightly supercritical Galton-Watson process. As Haldane concluded, the probability π of ultimate survival obeys the asymptotics

$$\pi \approx \frac{2\varepsilon}{\sigma^2}$$

for small $\varepsilon > 0$, where $1 + \varepsilon$ is the offspring expectation and σ^2 the offspring variance. This approximation gained a lot of attention in the literature, see [PW08] for an overview. For Galton-Watson processes this asymptotics was considered among others by Kolmogorov [Kol38], Eshel [Esh81], Athreya [Ath92] and Hoppe [Hop92].

In this paper we consider the same problem, and study the asymptotic survival probability of a slightly supercritical branching process in a random environment. As it turns out Haldane's asymptotics remains valid for branching processes in an

iid random environment, as long as one keeps away from subcritical behaviour. This might come as a surprise, since the corresponding Kolmogorov asymptotics for critical Galton-Watson processes does not turn up in the case of an iid random environment [GK00]. Close to subcriticality one observes a slightly different behaviour of the survival probability, a smooth adaption of Haldane's formula.

Let us recall the notion of a branching process in random environment. Denote by $\mathcal{P}(\mathbb{N}_0)$ the space of all probability measures on $\mathbb{N}_0 = \{0, 1, 2, \dots\}$. For $f \in \mathcal{P}(\mathbb{N}_0)$ it is convenient to identify the measure f with its probability generating function

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

where $f[z]$ denotes the weight of f in $z \in \mathbb{N}_0$. The mean and the second factorial moment can be expressed through the generating functions as

$$f'(1) = \sum_{z=1}^{\infty} z f[z], \quad f''(1) = \sum_{z=2}^{\infty} z(z-1) f[z].$$

Endow the space of probability measures $\mathcal{P}(\mathbb{N}_0)$ with the total variation metric and the induced Borel- σ -algebra. This allows to consider random probability measures on $\mathcal{P}(\mathbb{N}_0)$ and we define a random environment $V = (F_1, F_2, \dots)$ as a sequence of random probability measures. In this paper, we consider branching process in an iid random environment, in this case the environment $V = (F_1, F_2, \dots)$ consists of independent copies of a random measure F . As before we can express the now random mean and random second factorial moment as

$$F'(1) = \sum_{z=1}^{\infty} z F[z], \quad F''(1) = \sum_{z=2}^{\infty} z(z-1) F[z].$$

Given a sequence $V = (F_1, F_2, \dots)$ of random probability measures we call $(Z_n, n \geq 0)$ a branching process in random environment, if it has the representation

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{i,n}, \quad Z_0 = 1,$$

where conditionally on V the family $(\xi_{i,n}, i \geq 1, n \geq 1)$ is independent and for each $n \in \mathbb{N}$ the sequence $(\xi_{i,n}, i \geq 1)$ is identically distributed with distribution F_n . Unconditionally, the random variables $\xi_{i,n}$ are just identically distributed, thus copies of a generic random variable ξ . For more background on branching processes in random environment, see the monograph [KV17].

In the following we focus on supercritical branching processes in an iid random environment. It is known that they exhibit similar behaviour as supercritical Galton-Watson processes. Tanny [Tan88] derived the Kesten-Stigum theorem in the random environment setup, see also [Ham92]. More recently finer asymptotics have been achieved, among others in [BB09],[BB14], [GLM17] and [HL14].

Here, we consider a sequence $(Z^{(N)}, N \geq 1)$ of branching processes in an iid random environment and study the asymptotic behaviour of their survival probabilities. In order to keep the notation accessible we suppress the index N at random quantities and quote it at probabilities and expectations. Thus we write e.g. the

survival probability as

$$\pi_N := \lim_{n \rightarrow \infty} \mathbb{P}_N(Z_n > 0),$$

and not as $\lim_{n \rightarrow \infty} \mathbb{P}(Z_n^{(N)} > 0)$.

Let

$$\varepsilon_N := \mathbb{E}_N[F'(1)] - 1, \quad \nu_N := \text{Var}_N(F'(1)).$$

We are considering supercritical processes, implying $\varepsilon_N > 0$ for all N . Like Haldane, we are concerned with the situation $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$. We point out that even in this situation critical/subcritical behaviour is not excluded. However, if we stay in the supercritical regime, Haldane's asymptotics remarkably enough remains valid, apart from a narrow region of transition, where the processes behaviour changes from supercritical to subcritical. This region occurs if the standard deviation of $F'(1)$ is of order $\sqrt{\varepsilon_N}$, thus of notably larger order than ε_N , the expected excess of $F'(1)$ above 1. Only here the random environment gains impact.

Let us state the conditions of our main result. We assume that the annealed variance of ξ stabilizes for large N ,

$$\text{Var}_N(\xi) = \sigma^2 + o(1), \quad \mathbb{E}_N[\xi^4] = O(1) \quad \text{as } N \rightarrow \infty, \quad (5.1)$$

with $\sigma^2 > 0$. Without further mentioning we require $F'(1) > 0$ \mathbb{P}_N -almost surely, since otherwise $\pi_N = 0$ trivially. Moreover, we assume that for some $\delta > 0$

$$\mathbb{E}_N[F'(1)^{-4-\delta}] = O(1), \quad \mathbb{E}_N[|F'(1) - \mathbb{E}_N[F'(1)]|^{4+\delta}] = O(\nu_N^{2+\frac{\delta}{2}}). \quad (5.2)$$

Theorem 10. *Let $\varepsilon_N \rightarrow 0$ as $N \rightarrow \infty$. Then, under the assumptions (5.1) and (5.2), we have*

i) *If $\nu_N = o(\varepsilon_N)$, then the survival probability obeys*

$$\pi_N \sim \frac{2\varepsilon_N}{\sigma^2} \quad \text{as } N \rightarrow \infty.$$

ii) *If $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ with $0 < \rho < 2$, then*

$$\pi_N \sim \frac{(2-\rho)\varepsilon_N}{\sigma^2} \quad \text{as } N \rightarrow \infty.$$

iii) *If $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ with $2 < \rho < \infty$, then for large N*

$$\pi_N = 0.$$

In the case $\rho = 2$ a finer analysis of the asymptotics would be needed.

The proof of Theorem 10 follows a new strategy and does not rely on the familiar fixed point characterization of the survival probability. Instead we use an explicit representation for the survival probability, given in Section 5.2. It has a similar structure as perpetuities, which are known from a financial context. Indeed a major step of the proof is to derive a limit theorem for perpetuities with asymptotically vanishing interest rates. This result, given in Section 5.3, might be of independent interest. With these ingredients the proof of Theorem 10 is completed in Section 5.4.

5.2 An expression for the survival probability

In the following we derive an expression for the survival probability

$$\pi = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n > 0)$$

of a branching process in random environment $(Z_n, n \geq 0)$, using ideas of [Ker20]. Consider an iid sequence of random measures $V = (F_1, F_2, \dots)$ on $\mathcal{P}(\mathbb{N}_0)$. If we condition on the event $\{F_1 = f_1, F_2 = f_2, \dots\}$ for a sequence (f_1, f_2, \dots) in $\mathcal{P}(\mathbb{N}_0)$, we may express the conditional generating function of Z_n as

$$\mathbb{E}[s^{Z_n} \mid Z_0, \dots, Z_{n-1}, F_1 = f_1, \dots, F_n = f_n] = f_n(s)^{Z_{n-1}}, \quad 0 \leq s \leq 1.$$

Iterating this formula yields

$$\mathbb{E}[s^{Z_n} \mid F_1 = f_1, \dots, F_n = f_n] = f_1(f_2(\dots f_n(s) \dots)), \quad 0 \leq s \leq 1.$$

We set for $k \leq n - 1$

$$g_{k,n}(s) := f_{k+1} \circ \dots \circ f_n(s), \quad 0 \leq s \leq 1,$$

with the convention $g_{n,n}(s) = s$. Similarly to classical Galton-Watson processes we obtain

$$\mathbb{P}(Z_n > 0 \mid F_1 = f_1, \dots, F_n = f_n) = 1 - g_{0,n}(0). \quad (5.3)$$

Next, for any probability generating function f on \mathbb{N} with positive finite mean m as in [Ker20] we define the so-called shape function $\psi : [0, 1) \rightarrow \mathbb{R}$ via the equation

$$\frac{1}{1 - f(s)} = \frac{1}{m(1 - s)} + \psi(s).$$

ψ can be extended continuously to $[0, 1]$ by setting

$$\psi(1) := \frac{f''(1)}{2f'(1)^2}.$$

Noting that $g_{k,n}(s)$ is also a probability generating function and letting ψ_k be shape the function belonging to f_k , we obtain iteratively

$$\begin{aligned} \frac{1}{1 - g_{0,n}(s)} &= \frac{1}{1 - f_1(g_{1,n}(s))} = \frac{1}{f'_1(1)(1 - g_{1,n}(s))} + \psi_1(g_{1,n}(s)) \\ &= \frac{1}{f'_1(1)f'_2(1)(1 - g_{2,n}(s))} + \frac{\psi_2(g_{2,n}(s))}{f'_1(1)} + \psi_1(g_{1,n}(s)) \\ &= \frac{1}{f'_1(1) \cdots f'_n(1)(1 - s)} + \sum_{k=0}^{n-1} \frac{\psi_{k+1}(g_{k+1,n}(s))}{f'_1(1) \cdots f'_k(1)}. \end{aligned}$$

In the unconditional setting, we write $\varphi_k(s)$ for the shape function of $F_k(s)$, in particular

$$\varphi_k(1) = \frac{F''_k(1)}{2F'_k(1)^2}, \quad k \in \mathbb{N} \quad (5.4)$$

and let

$$\mu_k = \prod_{i=1}^k F'_i(1), \quad \mu_0 = 1.$$

Using (5.3) and (5.4) we obtain

$$\frac{1}{\mathbb{P}(Z_n > 0 \mid V)} = \frac{1}{\mu_n} + \sum_{k=0}^{n-1} \frac{\varphi_{k+1}(\mathbb{P}(Z_n = 0 \mid Z_{k+1} = 1, V))}{\mu_k} \quad \text{a.s.} \quad (5.5)$$

Equation (5.5) will be crucial in the proof of the following Proposition.

Proposition 5.2.1. *Let $(Z_n, n \geq 0)$ be a branching process in random environment $V = (F_1, F_2, \dots)$ consisting of independent copies of F . Assume $0 < \mathbb{E}[\log F'(1)] < \infty$ and $\mathbb{E}[\log^+ F''(1)] < \infty$. Then the survival probability π can be expressed as*

$$\pi = \mathbb{E} \left[\frac{1}{X} \right], \quad (5.6)$$

with

$$X := \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(\mathbb{P}(Z_{\infty} = 0 \mid Z_{k+1} = 1, V))}{\mu_k} < \infty \quad \text{a.s.}$$

Remark 5.2.2. *The assumptions in Proposition 5.2.1 are slightly stronger than the classical requirements for supercriticality, see [SW69] and [Tan77]. However, these references contain no representation as (5.6). We note that Proposition 5.2.1 holds equally for a stationary and ergodic random environment.*

Proof. Taking the limit $n \rightarrow \infty$ in (5.5) yields, with $Z_{\infty} := \lim_{n \rightarrow \infty} Z_n$

$$\frac{1}{\mathbb{P}(Z_{\infty} > 0 \mid V)} = \lim_{n \rightarrow \infty} \left[\frac{1}{\mu_n} + \sum_{k=0}^{n-1} \frac{\varphi_{k+1}(\mathbb{P}(Z_n = 0 \mid Z_{k+1} = 1, V))}{\mu_k} \right].$$

Observe that $\mu_n \rightarrow \infty$ almost surely, since

$$\mu_n = \exp \left(\sum_{j=1}^n \log F'_j(1) \right) = \exp \left(n \mathbb{E}[\log F'(1)] + o(n) \right), \quad (5.7)$$

by the strong law of large numbers. Thus

$$\frac{1}{\mathbb{P}(Z_{\infty} > 0 \mid V)} = \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \frac{\varphi_{k+1}(\mathbb{P}(Z_n = 0 \mid Z_{k+1} = 1, V))}{\mu_k}.$$

Furthermore, we have $\mathbb{P}(Z_n = 0 \mid Z_{k+1} = 1, V) \rightarrow \mathbb{P}(Z_{\infty} = 0 \mid Z_{k+1} = 1, V)$ a.s. which yields

$$\varphi_{k+1}(\mathbb{P}(Z_n = 0 \mid Z_{k+1} = 1, V)) \rightarrow \varphi_{k+1}(\mathbb{P}(Z_{\infty} = 0 \mid Z_{k+1} = 1, V)) \quad \text{a.s.}$$

by continuity. This implies

$$\frac{1}{\mathbb{P}(Z_\infty > 0 \mid V)} = \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(\mathbb{P}(Z_\infty = 0 \mid Z_{k+1} = 1, V))}{\mu_k}, \quad (5.8)$$

provided that we can justify the interchange of limits. By Lemma 1 in [Ker20] we have for $k \leq n$

$$\frac{\varphi_{k+1}(\mathbb{P}(Z_n = 0 \mid Z_{k+1} = 1, V))}{\mu_k} \leq 2 \frac{\varphi_{k+1}(1)}{\mu_k}.$$

Thus, letting

$$Y := \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(1)}{\mu_k},$$

it is sufficient in view of dominated convergence to prove $Y < \infty$ almost surely. Note that $\varphi_k(1) = \exp(\log F_k''(1) - 2 \log F_k'(1))$ and by the strong law of large numbers and a known Borel-Cantelli argument we have $\log F_k'(1) = o(k)$ and $\log^+ F_k''(1) = o(k)$, therefore $\varphi_k(1) = e^{o(k)}$. Together with (5.7) it follows $Y < \infty$ almost surely. \square

5.3 On perpetuities with small interest rates

Let $(A_n, B_n), n \geq 1$, be independent copies of the random pair (A, B) and assume that the components A, B are nonnegative and independent. Define $C_n := B_1 B_2 \cdots B_n$, $C_0 = 1$ for $n \in \mathbb{N}$ and consider the series

$$Y = \sum_{k=0}^{\infty} C_k A_{k+1},$$

which in a financial context is called a perpetuity, see [AIR09] and the literature cited therein. The random variable Y fulfils a stochastic recursion, the *annuity equation* [BDM16]

$$Y \stackrel{d}{=} A + BY, \quad (5.9)$$

where on the right hand side (A, B) and Y are assumed to be independent.

In the following we study the limiting behaviour of a sequence of perpetuities Y as the expectation of B tends to 1 (which in a financial setting corresponds to vanishing interest rates). As before we keep the sequence index N just as a subscript of the expectations and probabilities. Let

$$\beta_N := 1 - \mathbb{E}_N[B], \quad \gamma_N := \text{Var}_N(B).$$

We stay in the region $\beta_N > -\frac{\gamma_N}{2}$ (elsewhere one typically has $Y = \infty$ almost surely). Depending on the ratio of $\frac{\beta_N}{\gamma_N}$ we observe different limiting distributions for rescaled versions of Y .

Theorem 11. *Assume in the limit $N \rightarrow \infty$*

$$\beta_N \rightarrow 0, \quad \gamma_N \rightarrow 0, \quad \mathbb{E}_N[A] = \alpha + o(1), \quad \mathbb{E}_N[A^{1+\delta}] = O(1),$$

with $\alpha > 0$ and some $\delta > 0$, furthermore $\mathbb{E}_N [|B - 1|^{2+\delta}] = o(\beta_N + \gamma_N)$. Additionally assume

$$\frac{\beta_N}{\gamma_N} \rightarrow \hat{\rho} \quad \text{with} \quad -\frac{1}{2} < \hat{\rho} \leq \infty.$$

Then we have:

- i) If $\hat{\rho} = \infty$, then $\beta_N Y$ converges to α in probability.
- ii) If $\hat{\rho} \in (-1/2, \infty)$, then $\gamma_N Y$ is asymptotically inverse gamma distributed, with density $\frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x} dx$ on \mathbb{R}^+ and parameters $(a, b) = (2\hat{\rho} + 1, 2\alpha)$.

We prepare the proof of Theorem 11 by three lemmata.

Lemma 5.3.1. Let τ_N denote either β_N or γ_N , and define

$$\ell_\tau(\lambda) = \mathbb{E}_N [\exp(-\lambda \tau_N Y)]$$

to be the Laplace transform of $\tau_N Y$. Then under the assumptions of Theorem 11, for any $\lambda > 0$,

- i) $\mathbb{E}_N [e^{-\lambda \tau_N A}] = 1 - \lambda \tau_N \alpha + o(\tau_N)$,
- ii) $\mathbb{E}_N [\ell_\tau(\lambda B)] = \ell_\tau(\lambda) - \beta_N \lambda \ell'_\tau(\lambda) + \frac{\gamma_N}{2} \lambda^2 \ell''_\tau(\lambda) + o(\beta_N + \gamma_N)$.

Proof. i): We split the expectation $\mathbb{E}_N [e^{-\lambda \tau_N A}]$ into two parts. On the one hand we obtain for any $\eta > 0$

$$\mathbb{E}_N [e^{-\lambda \tau_N A}; \tau_N A > \eta] \leq \mathbb{P}_N (\tau_N A > \eta) \leq \frac{\tau_N^{1+\delta}}{\eta^{1+\delta}} \mathbb{E}_N [A^{1+\delta}] = o(\tau_N). \quad (5.10)$$

On the other hand, using the Taylor expansion $e^{-\lambda \tau_N A} = 1 - \lambda \tau_N A e^{-\lambda \tau_N A'}$ with $0 \leq A' \leq A$, we have

$$\begin{aligned} & \left| \mathbb{E}_N [e^{-\lambda \tau_N A}; \tau_N A \leq \eta] - \mathbb{E}_N [1 - \lambda \tau_N A] \right| \\ & \leq \left| \mathbb{E}_N [1 - \lambda \tau_N A; \tau_N A > \eta] \right| + \left| \mathbb{E}_N [\lambda \tau_N A (1 - e^{-\lambda \tau_N A'}); \tau_N A \leq \eta] \right| \\ & \leq \mathbb{P}_N (\tau_N A > \eta) + \mathbb{E}_N [\lambda \tau_N A; \tau_N A > \eta] + \mathbb{E}_N [\lambda \tau_N A (1 - e^{-\lambda \eta})] \\ & \leq \left(\frac{\tau_N^{1+\delta}}{\eta^{1+\delta}} + \frac{\lambda \tau_N^{1+\delta}}{\eta^\delta} \right) \mathbb{E}_N [A^{1+\delta}] + \mathbb{E}_N [\lambda \tau_N A (1 - e^{-\lambda \eta})] \\ & = \lambda \tau_N \mathbb{E}_N [A] (1 - e^{-\lambda \eta}) + O(\tau_N^{1+\delta}). \end{aligned}$$

Therefore, together with (5.10) we obtain for any $\eta > 0$

$$\left| \mathbb{E}_N [e^{-\lambda \tau_N A}] - (1 - \lambda \tau_N \alpha) \right| \leq \lambda \tau_N \alpha (1 - e^{-\lambda \eta}) + o(\tau_N).$$

Since the left-hand side does not depend on η and η can be chosen arbitrarily small, we have proven i).

ii): Again we split the expectation $\mathbb{E}_N [\ell_\tau(B\lambda)]$ into two parts. We have for any $\eta > 0$

$$\mathbb{E}_N [\ell_\tau(B\lambda); |B - 1| > \eta] \leq \mathbb{P}_N (|B - 1| > \eta) \leq \frac{\mathbb{E}_N [|B - 1|^{2+\delta}]}{\eta^{2+\delta}} = o(\beta_N + \gamma_N).$$

Using a Taylor expansion we get for some B' between 1 and B

$$\ell_\tau(\lambda B) = \ell_\tau(\lambda) + \ell'_\tau(\lambda)(\lambda B - \lambda) + \frac{1}{2}\ell''_\tau(\lambda B')(\lambda - \lambda B)^2.$$

Therefore,

$$\begin{aligned} & \left| \mathbb{E}_N [\ell_\tau(\lambda B); |B - 1| \leq \eta] - \mathbb{E}_N \left[\ell_\tau(\lambda) + \ell'_\tau(\lambda)(\lambda B - \lambda) + \frac{1}{2}\ell''_\tau(\lambda)(\lambda B - \lambda)^2 \right] \right| \\ & \leq \left| \mathbb{E}_N \left[\ell_\tau(\lambda) + \ell'_\tau(\lambda)(\lambda B - \lambda) + \frac{1}{2}\ell''_\tau(\lambda)(\lambda B - \lambda)^2; |B - 1| > \eta \right] \right| \\ & \quad + \left| \mathbb{E}_N [(\ell''_\tau(\lambda) - \ell''_\tau(\lambda B'))(\lambda B - \lambda)^2; |B - 1| \leq \eta] \right| \\ & = \left| \mathbb{E}_N [\ell'''_\tau(\lambda B'')\lambda(B' - 1)(\lambda B - \lambda)^2; |B - 1| \leq \eta] \right| + O(\mathbb{E}_N [|B - 1|^{2+\delta}]) \\ & \leq \eta\lambda^3 |\ell'''_\tau(\lambda(1 - \eta))| \mathbb{E}_N [(B - 1)^2; |B - 1| \leq \eta] + o(\beta_N + \gamma_N), \end{aligned} \quad (5.11)$$

with some B'' between B' and 1. We have the estimate

$$|\ell'''_\tau(\lambda(1 - \eta))| = \mathbb{E}_N [(\tau_N Y)^3 e^{-\lambda(1-\eta)\tau_N Y}] \leq \sup_{y \in \mathbb{R}^+} y^3 e^{-\lambda(1-\eta)y} < \infty.$$

This allows to bound (5.11) from above for some constant $c > 0$ by the expression

$$c\eta\lambda^3 \mathbb{E}_N [\lambda\eta(B - 1)^2; |B - 1| \leq \eta] \leq c\eta\lambda^3(\beta_N^2 + \gamma_N).$$

Therefore, putting our estimates together, we have shown that for any $\eta > 0$

$$\begin{aligned} & \left| \mathbb{E}_N [\ell_\tau(\lambda B)] - \mathbb{E}_N \left[\ell_\tau(\lambda) + \ell'_\tau(\lambda)(\lambda B - \lambda) + \frac{1}{2}\ell''_\tau(\lambda)(\lambda B - \lambda)^2 \right] \right| \\ & \leq c\eta\lambda^3(\beta_N + \gamma_N) + o(\beta_N + \gamma_N). \end{aligned}$$

Letting $\eta \rightarrow 0$, the claim ii) follows. \square

Lemma 5.3.2. *Under the assumptions of Theorem 11, for all $\eta > 0$ there exists a constant $c = c(\eta, \hat{\rho}) > 0$ (not depending on $\mathbb{E}_N [A]$) such that for $\beta_N + \gamma_N$ sufficiently small*

$$\mathbb{P}_N \left((\beta_N + \gamma_N) \sum_{k=0}^{\infty} C_k A_{k+1} \geq c\mathbb{E}_N [A] \right) \leq \eta.$$

Proof. First we determine an $u \in (0, 1)$ such that $\mathbb{E}_N [B^u] < 1$. Hence, we split $\mathbb{E}_N [B^u]$ into two parts and obtain

$$\begin{aligned} \mathbb{E}_N [B^u; |B - 1| > u] & \leq \mathbb{P}_N (B - 1 < -u) + \frac{(1 + u)^u}{u^u} \mathbb{E}_N [(B - 1)^u; B - 1 > u] \\ & \leq \mathbb{E}_N [|B - 1|^{2+\delta}] \left(\frac{1}{u^{2+\delta}} + \frac{(1 + u)^u}{u^{2+\delta}} \right) \\ & = o(\beta_N + \gamma_N). \end{aligned}$$

Moreover, by means of a Taylor expansion

$$\begin{aligned}
\mathbb{E}_N [B^u; |B-1| \leq u] &\leq \mathbb{E}_N \left[1 + u(B-1) + \frac{u(u-1)}{2}(1+u)^{u-2}(B-1)^2; |B-1| \leq u \right] \\
&\leq 1 - u\beta_N - \frac{u(1-u)}{2(1+u)^{2-u}}(\gamma_N + \beta_N^2) \\
&\quad - u\mathbb{E}_N \left[(B-1) + \frac{u-1}{2(1+u)^{2-u}}(B-1)^2; |B-1| > u \right] \\
&= 1 - u(\beta_N + \frac{1}{2}\frac{1-u}{(1+u)^{2-u}}(\gamma_N + \beta_N^2)) + O\left(\mathbb{E}_N [|B-1|^{2+\delta}]\right) \\
&= 1 - u(\beta_N + \frac{1}{2}\frac{1-u}{(1+u)^{2-u}}(\gamma_N + \beta_N^2)) + o(\beta_N + \gamma_N).
\end{aligned}$$

Due to our assumption $\frac{\beta_N}{\gamma_N} \rightarrow \hat{\rho} \in (-\frac{1}{2}, \infty]$ there exists $u = u_{\hat{\rho}} \in (0, 1)$, such that $\beta_N + \frac{1}{2}\frac{1-u}{(1+u)^{2-u}}(\gamma_N + \beta_N^2) + o(\beta_N + \gamma_N) \geq \theta(\beta_N + \gamma_N)$ for some $\theta = \theta_{\hat{\rho}} > 0$ and $\beta_N + \gamma_N$ small enough. This yields

$$\mathbb{E}_N [B^u] \leq 1 - \theta(\beta_N + \gamma_N) \leq e^{-\theta(\beta_N + \gamma_N)}. \quad (5.12)$$

Next, the process $M_k := \prod_{i=1}^k B_i^u / \mathbb{E}_N [B^u]$ is a nonnegative martingale, with $M_0 = 1$. For the stopping time $T = \min\{k \geq 0 : M_k \geq a\}$ $a \geq 1$, we get by means of the optional stopping theorem for all $t \in \mathbb{N}_0$

$$1 = \mathbb{E}_N [M_{T \wedge t}] \geq a\mathbb{P}_N (T \leq t).$$

Taking the limit $t \rightarrow \infty$ we obtain

$$\mathbb{P}_N (T < \infty) \leq \frac{1}{a}.$$

Together with (5.12) we conclude that on the event $\{T = \infty\}$ for all $k \geq 0$

$$\prod_{i=1}^k B_i = \left(\prod_{i=1}^k B_i^u \right)^{\frac{1}{u}} \leq \left(a \prod_{i=1}^k \mathbb{E}_N [B_i^u] \right)^{\frac{1}{u}} \leq a^{\frac{1}{u}} \exp \left(-\frac{\theta(\beta_N + \gamma_N)k}{u} \right).$$

This yields

$$\begin{aligned}
& \mathbb{P}_N \left((\beta_N + \gamma_N) \sum_{k=0}^{\infty} C_k A_{k+1} \geq c \mathbb{E}_N [A] \right) \\
& \leq \frac{1}{a} + \mathbb{P}_N \left((\beta_N + \gamma_N) a^{\frac{1}{u}} \sum_{k=0}^{\infty} \exp \left(-\frac{\theta(\beta_N + \gamma_N)k}{u} \right) A_{k+1} \geq c \mathbb{E}_N [A] \right) \\
& \leq \frac{1}{a} + \frac{(\beta_N + \gamma_N) a^{\frac{1}{u}}}{c \mathbb{E}_N [A]} \mathbb{E}_N \left[\sum_{k=0}^{\infty} \exp \left(-\frac{\theta(\beta_N + \gamma_N)k}{u} \right) A_{k+1} \right] \\
& = \frac{1}{a} + \frac{(\beta_N + \gamma_N) a^{\frac{1}{u}}}{c} \frac{1}{1 - \exp \left(-\frac{\theta(\beta_N + \gamma_N)}{u} \right)} \\
& \leq \frac{1}{a} + 2 \frac{a^{\frac{1}{u}} u}{c \theta}
\end{aligned}$$

for $\beta_N + \gamma_N$ small enough. Letting $a = \frac{2}{\eta}$ and $c = c(\eta, \hat{\rho})$ large enough, the probability above can be made smaller than any $\eta > 0$. \square

Remark 5.3.3. As we have seen in the proof of Lemma 5.3.2, it is only required that the random variables $A_k, k \geq 1$ are nonnegative and have one and the same expectation. This will be useful later in the proof of Lemma 5.4.5.

Lemma 5.3.4. Let W be inverse gamma distributed with parameters $a > 0, b > 0$, that is W has density $\frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x} dx$. The Laplace transform $h(\lambda) = \mathbb{E}[\exp(-\lambda W)]$ then fulfils the second order differential equation

$$\lambda h''(\lambda) = (a - 1)h'(\lambda) + bh(\lambda), \quad h(0) = 1.$$

Proof. By the definition of h we have

$$h(\lambda) = \frac{b^a}{\Gamma(a)} \int_0^{\infty} e^{-\lambda x} x^{-a-1} e^{-b/x} dx$$

Differentiating two times with respect to λ , gives

$$\begin{aligned}
h'(\lambda) &= -\frac{b^a}{\Gamma(a)} \int_0^{\infty} e^{-\lambda x} x^{-a} e^{-b/x} dx \\
h''(\lambda) &= \frac{b^a}{\Gamma(a)} \int_0^{\infty} e^{-\lambda x} x^{-a+1} e^{-b/x} dx.
\end{aligned}$$

Using integration by parts applied to the functions $-e^{-\lambda x}$ and $x^{-a+1} e^{-b/x}$ yields

$$\begin{aligned}
\lambda h''(\lambda) &= \frac{b^a}{\Gamma(a)} \left[[-e^{-\lambda x} x^{-a+1} e^{-b/x}]_0^{\infty} - \int_0^{\infty} e^{-\lambda x} \left((a-1)x^{-a} e^{-b/x} - bx^{-a-1} e^{-b/x} \right) dx \right] \\
&= -\frac{b^a}{\Gamma(a)} (a-1) \int_0^{\infty} e^{-\lambda x} x^{-a} e^{-b/x} dx + b \frac{b^a}{\Gamma(a)} \int_0^{\infty} e^{-\lambda x} x^{-a-1} e^{-b/x} dx \\
&= (a-1)h'(\lambda) + bh(\lambda),
\end{aligned}$$

which is the claimed result. \square

Proof of Theorem 11. The proofs of parts i) and ii) are using the same ideas. We will prove convergence of the Laplace transform of $\beta_N Y$ or $\gamma_N Y$ towards the Laplace transform of a Dirac distributed random variable or to the Laplace transform of an inverse gamma distributed random variable.

i): Recall the notation $\ell_\beta(\lambda) = \mathbb{E}_N[\exp(-\lambda\beta_N Y)]$ from Lemma 5.3.1. From (5.9) we have for any $\lambda > 0$

$$\ell_\beta(\lambda) = \mathbb{E}_N \left[e^{-\lambda\beta_N A} \right] \mathbb{E}_N \left[e^{-\lambda\beta_N B Y} \right] = \mathbb{E}_N \left[e^{-\lambda\beta_N A} \right] \mathbb{E}_N [\ell_\beta(\lambda B)],$$

by independence of A, B and Y . Since $\gamma_N = o(\beta_N)$, it follows by Lemma 5.3.1

$$\begin{aligned} \ell_\beta(\lambda) &= [1 - \beta_N \lambda \alpha + o(\beta_N)] [\ell_\beta(\lambda) - \beta_N \lambda \ell'_\beta(\lambda) + O(\gamma_N) + o(\beta_N)] \\ &= (1 - \beta_N \lambda \alpha) \ell_\beta(\lambda) - \beta_N \lambda \ell'_\beta(\lambda) + o(\beta_N) \end{aligned} \quad (5.13)$$

Due to Lemma 5.3.2 (or alternatively due to the fact that the annuity equation (5.9) implies $\mathbb{E}_N[\beta_N Y] \rightarrow \alpha$ as $\beta_N \rightarrow 0$), the family $\{\beta_N Y\}$ is tight. Moreover, let $\beta_{N'} Y$ be a subsequence converging in distribution. We have to show that the limiting distribution of $\beta_{N'} Y$ is unique. Now, convergence of $\beta_{N'} Y$ implies convergence of the Laplace-transforms together with their derivatives. Dividing both sides in (5.13) by $\lambda\beta_{N'}$ we obtain the limiting differential equation

$$0 = -\alpha \ell(\lambda) - \ell'(\lambda), \quad \ell(0) = 1,$$

which has the unique solution $\ell(\lambda) = e^{-\alpha\lambda}$, the Laplace transform of a random variable which takes the value α with probability 1. This proves part i) of Theorem 10.

ii): Choosing $\tau_N = \gamma_N$ in Lemma 5.3.1 we get by the same line of arguments as before

$$\begin{aligned} \ell_\gamma(\lambda) &= [1 - \gamma_N \lambda \alpha + o(\gamma_N)] \left[\ell_\gamma(\lambda) - \beta_N \lambda \ell'_\gamma(\lambda) + \frac{\gamma_N}{2} \lambda^2 \ell''_\gamma(\lambda) + o(\gamma_N) \right] \\ &= (1 - \gamma_N \lambda \alpha) \ell_\gamma(\lambda) - \beta_N \lambda \ell'_\gamma(\lambda) + \frac{\gamma_N}{2} \lambda^2 \ell''_\gamma(\lambda) + o(\gamma_N). \end{aligned} \quad (5.14)$$

Due to Lemma 5.3.2 the family $\{\gamma_N Y\}$ is tight. Moreover, let $\gamma_{N'} Y$ be a subsequence converging in distribution. Taking the limit and dividing both sides in (5.14) by $\lambda\gamma_{N'}$ we obtain the limiting differential equation

$$0 = -\alpha \ell(\lambda) - \hat{\rho} \ell'(\lambda) + \frac{1}{2} \lambda \ell''(\lambda), \quad \ell(0) = 1, \quad (5.15)$$

with $-\frac{1}{2} < \hat{\rho} < \infty$. This differential equation is a variant of the Bessel equation and can be solved by means of modified Bessel functions. However, we will not use this fact.

Instead we shall confirm that there is only one solution corresponding to a Laplace transform. To make this argument firm note that for any Laplace transform $\ell(\lambda)$ of a nonnegative random variable it holds that $\ell(\lambda) \rightarrow \mu_0$ as $\lambda \rightarrow \infty$, where μ_0 corresponds to the mass in 0 and that $\ell'(\lambda) \rightarrow 0$ as $\lambda \rightarrow \infty$. This implies, due to (5.15), that $\lambda \ell''(\lambda) \rightarrow 2\alpha\mu_0$ which in the case $\mu_0 > 0$ entails that $\ell'(\lambda) \rightarrow \infty$ and contradicts $\ell'(\lambda) \rightarrow 0$. Therefore, we have $\mu_0 = 0$ for any Laplace transform fulfilling (5.15).

Now take two Laplace transforms ℓ_1, ℓ_2 fulfilling (5.15). We show that $\ell_1 \equiv \ell_2$. Consider the difference $d = \ell_1 - \ell_2$. Since $d(0) = 0$ and $d(\lambda) \rightarrow 0$ as $\lambda \rightarrow \infty$ there

exists a $\lambda_0 > 0$ which is an extremal value, i.e. $d'(\lambda_0) = 0$. However, if $d(\lambda_0) > 0$ this implies due to (5.15) that $d''(\lambda_0) > 0$ which is a contradiction, and the same holds true, if $d(\lambda_0) < 0$. Therefore, we have $d \equiv 0$ and there is only one Laplace transform fulfilling (5.15). Lemma 5.3.4 shows, that $\gamma_N Y$ converges in distribution to an inverse gamma law with parameters $a = 2\hat{\rho} + 1$ and $b = 2\alpha$. \square

5.4 Proof of Theorem 10

We prepare the proof of Theorem 10 by several lemmata.

Lemma 5.4.1. *Under the assumptions of Theorem 10, we have*

- i) $\mathbb{E}_N [F'(1)^{-r}] = 1 - r\varepsilon_N + \frac{r(r+1)}{2}\nu_N + o(\varepsilon_N)$ for $r \in [0, 2]$,
- ii) $\mathbb{E}_N \left[\frac{F''(1)}{F'(1)^2} \right] = \sigma^2(1 + o(1))$.

Proof. i): Again we split the expectation $\mathbb{E}_N [F'(1)^{-r}]$ into parts. Let $0 < \eta < 1$, then

$$\begin{aligned} \mathbb{E}_N [F'(1)^{-r}; |F'(1) - 1| > \eta] &\leq \mathbb{E}_N [F'(1)^{-2r}]^{\frac{1}{2}} \mathbb{P}_N (|F'(1) - 1| > \eta)^{\frac{1}{2}} \\ &\leq \mathbb{E}_N [F'(1)^{-2r}]^{\frac{1}{2}} \frac{1}{\eta^{2+\delta/2}} \mathbb{E}_N [|F'(1) - 1|^{4+\delta}]^{\frac{1}{2}} \\ &= O(\nu_N^{1+\delta/2} + \varepsilon_N^{2+\delta}) = o(\varepsilon_N), \end{aligned} \quad (5.16)$$

due to assumption (5.2). A Taylor expansion for $F'(1)^{-r}$, yields with some U between $F'(1)$ and 1

$$F'(1)^{-r} = 1 - r(F'(1) - 1) + \frac{r(r+1)}{2}U^{-r-2}(F'(1) - 1)^2.$$

Therefore,

$$\begin{aligned} &\left| \mathbb{E}_N [F'(1)^{-r}; |F'(1) - 1| \leq \eta] - \mathbb{E}_N \left[1 - r(F'(1) - 1) + \frac{r(r+1)}{2}(F'(1) - 1)^2 \right] \right| \\ &\leq \left| \mathbb{E}_N \left[1 - r(F'(1) - 1) + \frac{r(r+1)}{2}(F'(1) - 1)^2; |F'(1) - 1| > \eta \right] \right| \\ &\quad + \left| \mathbb{E}_N \left[\frac{r(r+1)}{2}(F'(1) - 1)^2(1 - U^{-r-2}); |F'(1) - 1| \leq \eta \right] \right| \\ &\leq \frac{r(r+1)}{2} \mathbb{E}_N [(F'(1) - 1)^2] ((1 - \eta)^{-r-2} - 1) + O(\mathbb{E}_N [|F'(1) - 1|^{4+\delta}]) \\ &= ((1 - \eta)^{-r-2} - 1) (\nu_N + \varepsilon_N^2) + o(\varepsilon_N). \end{aligned}$$

Together with (5.16) we have shown

$$\left| \mathbb{E}_N [F'(1)^{-r}] - (1 - r\varepsilon_N + \frac{r(r+1)}{2}\nu_N) \right| \leq \frac{r(r+1)}{2} ((1 - \eta)^{-r-2} - 1) \nu_N + o(\varepsilon_N),$$

which proves i) by letting $\eta \rightarrow 0$.

ii): Due to assumption (5.1) we have

$$\begin{aligned}\mathbb{E}_N \left[\frac{F''(1)}{F'(1)^2} \right] &= \mathbb{E}_N [F''(1)] + \mathbb{E}_N \left[F''(1) \left(\frac{1}{F'(1)^2} - 1 \right) \right] \\ &= \sigma^2 + o(1) + \mathbb{E}_N \left[F''(1) \left(\frac{1}{F'(1)^2} - 1 \right) \right].\end{aligned}$$

Moreover,

$$\mathbb{E}_N \left[\left| F''(1) \left(\frac{1}{F'(1)^2} - 1 \right) \right| \right] \leq \mathbb{E}_N [F''(1)^2]^{\frac{1}{2}} \mathbb{E}_N \left[\left(\frac{1}{F'(1)^2} - 1 \right)^2 \right]^{\frac{1}{2}}.$$

The first term on the right-hand side is uniformly bounded by (5.1). For the second term note that the integrand converges to 0 in probability and is uniformly integrable due to (5.2), which implies that the second term converges to 0. This finishes the proof. \square

Lemma 5.4.2. *Under the assumptions of Theorem 10 there exists $\theta > 0$ and $p_0 \in (0, 1)$ such that for large N*

$$\mathbb{P}_N (\varphi(0) \geq \theta) \geq p_0.$$

Proof. By the definition of φ we have

$$\mathbb{P}_N (\varphi(0) \geq \theta) = \mathbb{P}_N \left(\frac{1}{1 - F(0)} \geq \theta + \frac{1}{F'(1)} \right).$$

From our assumptions we have $F'(1) \rightarrow 1$ in probability and therefore one obtains

$$\begin{aligned}\mathbb{P}_N (\varphi(0) \geq \theta) &\geq \mathbb{P}_N \left(\frac{1}{1 - F(0)} \geq 1 + 2\theta \right) + o(1) \\ &\geq \mathbb{P}_N (F(0) \geq 2\theta) + o(1) \\ &\geq \mathbb{E}_N [F(0)] - 2\theta + o(1).\end{aligned}\tag{5.17}$$

Thus, it suffices to bound $\mathbb{E}_N [F(0)]$ away from 0. Note that $\mathbb{E}_N [F(0)] = \mathbb{P}_N (\xi = 0)$. Since ξ^2 is uniformly integrable due to (5.1) we have for $k \geq 2$ and k large enough

$$\begin{aligned}\sigma^2 + o(1) &= \mathbb{E}_N [(\xi - 1)^2] = \mathbb{P}_N (\xi = 0) + \mathbb{E}_N [(\xi - 1)^2; 2 \leq \xi \leq k] + \mathbb{E}_N [(\xi - 1)^2; \xi > k] \\ &\leq \mathbb{P}_N (\xi = 0) + (k - 1)\mathbb{E}_N [\xi - 1; \xi \geq 2] + \frac{\sigma^2}{2} \\ &= \mathbb{P}_N (\xi = 0) + (k - 1)(\varepsilon_N + \mathbb{P}_N (\xi = 0)) + \frac{\sigma^2}{2}.\end{aligned}$$

This leads to the estimate

$$\frac{\sigma^2}{2k} + o(1) \leq \mathbb{P}_N (\xi = 0) = \mathbb{E}_N [F(0)].$$

Choosing θ small enough, (5.17) yields

$$\mathbb{P}_N (\varphi(0) \geq \theta) \geq \frac{\sigma^2}{2k} - 2\theta + o(1) \geq p_0,$$

for some $p_0 > 0$ and N sufficiently large. \square

We aim to study the limiting behaviour of the random variable X as $N \rightarrow \infty$. For this purpose we recall the notion

$$Y = \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(1)}{\mu_k},$$

with μ_k and $\varphi_k(1)$ as introduced in Section 5.2.

Lemma 5.4.3. *Under the assumptions of Theorem 10,*

- i) if $\frac{\nu_N}{\varepsilon_N} \rightarrow 0$ as $N \rightarrow \infty$, then $\varepsilon_N Y$ converges in probability to $\frac{\sigma^2}{2}$,
- ii) if $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ as $N \rightarrow \infty$ with $0 < \rho < 2$, then $\nu_N Y$ is asymptotically inverse gamma distributed with parameters $(a, b) = \left(\frac{2-\rho}{\rho}, \sigma^2\right)$.

Proof. The proof follows from Theorem 11. Note that Y fulfils the annuity equation (5.9),

$$Y \stackrel{d}{=} \varphi(1) + \frac{1}{F'(1)} Y,$$

with $A = \varphi(1)$ and $B = 1/F'(1)$. It remains to verify the assumptions from Theorem 11. By Lemma 5.4.1 ii)

$$\mathbb{E}_N [\varphi(1)] = \mathbb{E}_N \left[\frac{F''(1)}{2F'(1)^2} \right] = \frac{\sigma^2}{2} (1 + o(1)),$$

thus $\alpha = \frac{\sigma^2}{2}$. Also we have $\mathbb{E}_N [\varphi(1)^{1+\delta}] = O(1)$ for some $\delta > 0$ sufficiently small, due to

$$\begin{aligned} \mathbb{E}_N [\varphi(1)^{1+\delta}] &\leq \mathbb{E}_N \left[\frac{F''(1)^{1+\delta}}{F'(1)^{2+2\delta}} \right] \\ &\leq \mathbb{E}_N [F''(1)^{2+2\delta}]^{\frac{1}{2}} \mathbb{E}_N [F'(1)^{-4-4\delta}] = O(1) \end{aligned} \quad (5.18)$$

by assumption (5.1) and (5.2).

For $\beta_N = \mathbb{E}_N \left[\frac{1}{F'(1)} \right] - 1$, Lemma 5.4.1 i) yields

$$\beta_N = \varepsilon_N - \nu_N + o(\varepsilon_N).$$

Similarly, we obtain for $\gamma_N = \text{Var}_N (1/F'(1))$

$$\begin{aligned} \gamma_N &= \mathbb{E}_N [F'(1)^{-2}] - \mathbb{E}_N [F'(1)^{-1}]^2 \\ &= 1 - 2\varepsilon_N + 3\nu_N + o(\varepsilon_N) - (1 - \varepsilon_N + \nu_N + o(\varepsilon_N))^2 \\ &= \nu_N + o(\varepsilon_N). \end{aligned}$$

Finally, we have to confirm $\mathbb{E}_N \left[\left| \frac{1}{F'(1)} - 1 \right|^{2+\delta} \right] = o(\beta_N + \gamma_N) = o(\varepsilon_N)$. We have for $\delta > 0$ sufficiently small

$$\begin{aligned} \mathbb{E}_N \left[\left| \frac{1}{F'(1)} - 1 \right|^{2+\delta} \right] &= \mathbb{E}_N \left[\left| \frac{1 - F'(1)}{F'(1)} \right|^{2+\delta} \right] \\ &\leq \mathbb{E}_N \left[|1 - F'(1)|^{4+2\delta} \right]^{\frac{1}{2}} \mathbb{E}_N \left[F'(1)^{-4-2\delta} \right]^{\frac{1}{2}} \\ &= O(\nu_N^{\frac{2+\delta}{2}} + \varepsilon_N^{2+\delta}) = o(\varepsilon_N), \end{aligned}$$

by assumption (5.2).

Altogether we may apply Theorem 11. In case i) the assumption $\nu_N/\varepsilon_N \rightarrow 0$ implies

$$\frac{\gamma_N}{\beta_N} = \frac{\nu_N}{\varepsilon_N - \nu_N} \rightarrow 0,$$

hence $\varepsilon_N Y \rightarrow \alpha = \frac{\sigma^2}{2}$ in probability.

In case ii), that is $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ with $0 < \rho < 2$, we have

$$\frac{\beta_N}{\gamma_N} = \frac{\varepsilon_N - \nu_N + o(\varepsilon_N)}{\nu_N} \rightarrow \frac{1 - \rho}{\rho}.$$

Thus, by an application of Theorem 11 ii) $\nu_N Y$ is asymptotically inverse gamma distributed with parameters

$$(a, b) = \left(\frac{2(1 - \rho)}{\rho} + 1, \sigma^2 \right).$$

This concludes the proof. □

Recall from Proposition 5.2.1 the notation

$$X = \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(\mathbb{P}_N(Z_{\infty} = 0 \mid Z_{k+1} = 1, V))}{\mu_k}.$$

Lemma 5.4.4. *Under the conditions of Theorem 10 i) or ii), the sequence $\left(\frac{1}{\varepsilon_N X}, N \geq 1 \right)$ is uniformly integrable.*

Proof. We have to prove that for any choice of η there exists a constant $c > 0$ such that $\mathbb{E}_N \left[\frac{1}{\varepsilon_N X}; \frac{1}{\varepsilon_N X} > c \right] \leq \eta$. Since $X \geq 1$ almost surely due to (5.8), we have for any $\kappa_0 \in \mathbb{N}_0$

$$\begin{aligned} \mathbb{E}_N \left[\frac{1}{\varepsilon_N X}; \frac{1}{\varepsilon_N X} \geq 2^{\kappa_0} \right] &= \frac{1}{\varepsilon_N} \mathbb{E}_N \left[\frac{1}{X}; X \leq \frac{2^{-\kappa_0}}{\varepsilon_N} \right] \\ &\leq \frac{1}{\varepsilon_N} \sum_{\kappa=\kappa_0}^{\lfloor \frac{\log 1/\varepsilon_N}{\log 2} \rfloor} \varepsilon_N 2^{\kappa+1} \mathbb{P}_N \left(\frac{2^{-(\kappa+1)}}{\varepsilon_N} < X \leq \frac{2^{-\kappa}}{\varepsilon_N} \right) \\ &\leq \sum_{\kappa=\kappa_0}^{\lfloor \frac{\log 1/\varepsilon_N}{\log 2} \rfloor} 2^{\kappa+1} \mathbb{P}_N \left(X \leq \frac{2^{-\kappa}}{\varepsilon_N} \right). \end{aligned} \tag{5.19}$$

Let us estimate the right hand probabilities. Using Lemma 1 in [Ker20] we have for φ_{k+1} the lower bound $\frac{1}{2}\varphi_{k+1}(0)$ and get

$$\mathbb{P}_N \left(X \leq \frac{2^{-\kappa}}{\varepsilon_N} \right) \leq \mathbb{P}_N \left(\sum_{k=0}^{\infty} \frac{1}{2} \frac{\varphi_{k+1}(0)}{\mu_k} \leq \frac{2^{-\kappa}}{\varepsilon_N} \right) \leq \mathbb{P}_N \left(\sum_{k=0}^{\lfloor \varepsilon_N^{-1} \rfloor} \frac{\varphi_{k+1}(0)}{\mu_k} \leq \frac{2^{-\kappa+1}}{\varepsilon_N} \right).$$

Let $W = \#\{j \leq \lfloor \varepsilon_N^{-1} \rfloor : \varphi_j(0) \geq \theta\}$. Noting that $\varphi_0(0), \varphi_1(0), \dots$ are independent and identically distributed, the probability $\mathbb{P}_N \left(W < \frac{q}{\varepsilon_N} \right)$ is of order $o(\varepsilon_N)$ for any $q < p_0$ due to Lemma 5.4.2 and e.g. Hoeffding's inequality. Therefore,

$$\begin{aligned} \mathbb{P}_N \left(X \leq \frac{2^{-\kappa}}{\varepsilon_N} \right) &\leq \mathbb{P}_N \left(\sum_{k=0}^{\lfloor \varepsilon_N^{-1} \rfloor} \frac{\varphi_{k+1}(0)}{\mu_k} \leq \frac{2^{-\kappa+1}}{\varepsilon_N}, W \geq \frac{q}{\varepsilon_N} \right) + \mathbb{P}_N \left(W < \frac{q}{\varepsilon_N} \right) \\ &\leq \mathbb{P}_N \left(\frac{q\theta}{\varepsilon_N} \min_{k \leq \lfloor \varepsilon_N^{-1} \rfloor} \mu_k^{-1} \leq \frac{2^{-\kappa+1}}{\varepsilon_N} \right) + o(\varepsilon_N). \end{aligned}$$

Next, define the random variables $\zeta_j, j \geq 1$ by the equation $F'_j(1) = 1 + \varepsilon_N + \sqrt{\nu_N} \zeta_j$, thus one has $\mathbb{E}_N [\zeta_j] = 0$ and $\text{Var}_N (\zeta_j) = 1$. Also let $S_k = \sum_{j=1}^k \zeta_j$, then

$$\begin{aligned} \mu_k &= \prod_{j=1}^k (1 + \varepsilon_N + \sqrt{\nu_N} \zeta_j) \leq \exp(k\varepsilon_N + \sqrt{\nu_N} \sum_{j=1}^k \zeta_j) \\ &\leq \exp(1 + \sqrt{\nu_N} S_k), \end{aligned}$$

for $k \leq \lfloor \varepsilon_N^{-1} \rfloor$. This implies

$$\begin{aligned} \mathbb{P}_N \left(X \leq \frac{2^{-\kappa}}{\varepsilon_N} \right) &\leq \mathbb{P}_N \left(q\theta \min_{k \leq \lfloor \varepsilon_N^{-1} \rfloor} (\exp(1 + \sqrt{\nu_N} S_k))^{-1} \leq 2^{-\kappa+1} \right) + o(\varepsilon_N) \\ &\leq \mathbb{P}_N \left(\max_{k \leq \lfloor \varepsilon_N^{-1} \rfloor} S_k \geq \frac{1}{\sqrt{\nu_N}} \log \left(2^{\kappa-1} \frac{q\theta}{e} \right) \right) + o(\varepsilon_N) \end{aligned}$$

Note that the term $o(\varepsilon_N)$ above does not depend on κ . We aim to show that for some $c > 0$ and any κ such that $\log(2^{\kappa} \frac{q\theta}{2e}) > 0$

$$\mathbb{P}_N \left(\max_{k \leq \lfloor \varepsilon_N^{-1} \rfloor} S_k \geq \frac{\log \left(2^{\kappa} \frac{q\theta}{2e} \right)}{\sqrt{\nu_N}} \right) \leq \exp \left(-c \frac{\varepsilon_N (\log 2^{\kappa})^2}{\nu_N} \right) + o(\varepsilon_N), \quad (5.20)$$

with $o(\varepsilon_N)$ holding uniformly in κ . We defer the proof of (5.20) to the end. Applying the bound (5.20) and plugging this back into (5.19) we obtain

$$\begin{aligned} \mathbb{E}_N \left[\frac{1}{\varepsilon_N X}, \frac{1}{\varepsilon_N X} > 2^{\kappa_0} \right] &\leq \sum_{\kappa=\kappa_0}^{\lfloor \frac{\log 1/\varepsilon_N}{\log 2} \rfloor} 2^{\kappa+1} \left(\mathbb{P}_N \left(\max_{k \leq \lfloor \varepsilon_N^{-1} \rfloor} S_k \geq \frac{\left(\log 2^\kappa \frac{q\theta}{2e} \right)}{\sqrt{\nu_N}} \right) + o(\varepsilon_N) \right) \\ &\leq \sum_{\kappa=\kappa_0}^{\lfloor \frac{\log 1/\varepsilon_N}{\log 2} \rfloor} 2^{\kappa+1} \left(\exp \left[-c \frac{(\log 2^\kappa)^2 \varepsilon_N}{\nu_N} \right] + o(\varepsilon_N) \right) \\ &\leq \sum_{\kappa=\kappa_0}^{\infty} \left(2^{\kappa+1} \exp \left[-c \frac{\kappa^2 (\log 2)^2 \varepsilon_N}{\nu_N} \right] \right) + o(1). \end{aligned}$$

Due to the assumptions in Theorem 10 i) or ii) we have $\nu_N = O(\varepsilon_N)$. Also, the series $\sum_{\kappa} 2^\kappa e^{-c\kappa^2}$ is convergent for all $c > 0$, therefore

$$\mathbb{E}_N \left[\frac{1}{\varepsilon_N X}; \frac{1}{\varepsilon_N X} \geq 2^{\kappa_0} \right] \leq \eta + o(1),$$

if one chooses κ_0 big enough, which is the claimed result of the lemma.

It remains to prove (5.20), which we achieve via a Skorokhod embedding of $S_k = \sum_{i=1}^k \zeta_i$ into a standard Wiener process $(B_t)_{t \geq 0}$. The Skorokhod embedding theorem states that there exists a sequence of stopping times $0 = \tau_0 \leq \tau_1 \leq \tau_2, \dots$ such that

$$(S_1, S_2, \dots, S_k) \stackrel{d}{=} (B_{\tau_1}, B_{\tau_2}, \dots, B_{\tau_k}).$$

with $\mathbb{E}_N [\tau_i - \tau_{i-1}] = \mathbb{E}_N [\zeta_i^2] = 1$ and $\mathbb{E}_N [(\tau_i - \tau_{i-1})^{2+\delta}] \leq c_{2+\delta} \mathbb{E}_N [\zeta_i^{4+2\delta}] = O(1)$ due to assumption (5.2) for some positive constant $c_{2+\delta}$ only depending on δ , see [Roo69]. We estimate

$$\begin{aligned} &\mathbb{P}_N \left(\max_{k \leq \lfloor \varepsilon_N^{-1} \rfloor} S_k \geq \frac{\log \left(2^\kappa \frac{q\theta}{2e} \right)}{\sqrt{\nu_N}} \right) \\ &\leq \mathbb{P}_N \left(\sup_{t \leq \tau_{\lfloor 1/\varepsilon_N \rfloor}} B_t \geq \frac{\log \left(2^\kappa \frac{q\theta}{2e} \right)}{\sqrt{\nu_N}}, \tau_{\lfloor 1/\varepsilon_N \rfloor} \leq 2 \lfloor \varepsilon_N^{-1} \rfloor \right) + \mathbb{P}_N (\tau_{\lfloor 1/\varepsilon_N \rfloor} > 2 \lfloor \varepsilon_N^{-1} \rfloor) \\ &\leq \mathbb{P}_N \left(\sup_{t \leq t_0} B_t \geq \frac{\log \left(2^\kappa \frac{q\theta}{2e} \right)}{\sqrt{\nu_N}} \right) + \mathbb{P}_N (\tau_{\lfloor 1/\varepsilon_N \rfloor} > 2 \lfloor \varepsilon_N^{-1} \rfloor), \end{aligned} \tag{5.21}$$

with $t_0 = 2\varepsilon_N^{-1}$. The probability in the first term on the r.h.s of (5.21) simply concerns the maximum of a standard Brownian motion and can be estimated as

$$\mathbb{P}_N \left(\sup_{t \leq t_0} B_t \geq \frac{\log \left(2^\kappa \frac{q\theta}{2e} \right)}{\sqrt{\nu_N}} \right) = 2\mathbb{P}_N \left(B_{t_0} \geq \frac{\log \left(2^\kappa \frac{q\theta}{2e} \right)}{\sqrt{\nu_N}} \right) \leq 2 \exp \left(-c \frac{\varepsilon_N (\log 2^\kappa)^2}{\nu_N} \right),$$

for some $c > 0$. For the right-hand probability in (5.21) we express $\tau_{\lfloor 1/\varepsilon_N \rfloor}$ as a sum of independently and identically distributed random variables $(T_i, i \in \mathbb{N})$, such that

$$\tau_{\lfloor 1/\varepsilon_N \rfloor} = \sum_{i=1}^{\lfloor 1/\varepsilon_N \rfloor} T_i, \quad T_i := \tau_i - \tau_{i-1}.$$

We obtain for some constant c' only depending on δ ,

$$\begin{aligned} \mathbb{P}_N (\tau_{\lfloor 1/\varepsilon_N \rfloor} > 2\lfloor \varepsilon_N^{-1} \rfloor) &\leq \mathbb{P}_N \left(\sum_{i=1}^{\lfloor 1/\varepsilon_N \rfloor} (T_i - \mathbb{E}_N [T_i]) \geq \frac{1}{\varepsilon_N} \right) \\ &\leq (\varepsilon_N)^{2+\delta} \mathbb{E}_N \left[\left| \sum_{i=1}^{\lfloor 1/\varepsilon_N \rfloor} (T_i - \mathbb{E}_N [T_i]) \right|^{2+\delta} \right] \\ &\leq (\varepsilon_N)^{2+\delta} c' \left(\frac{1}{\varepsilon_N} \right)^{\frac{2+\delta}{2}} \\ &= o(\varepsilon_N). \end{aligned}$$

This finishes the proof. \square

Lemma 5.4.5. *Under the assumptions of Theorem 10, the random variable $\varepsilon_N(X - Y)$ converges to 0 in probability.*

Proof. Letting $Q_k = \mathbb{P}_N (Z_\infty = 0 \mid Z_k = 1, V)$ we have

$$\varepsilon_N |Y - X| \leq \varepsilon_N \sum_{k=0}^{\infty} \frac{1}{\mu_k} |\varphi_{k+1}(Q_{k+1}) - \varphi_{k+1}(1)|.$$

Thus, we prove that the right-hand side converges to 0 in probability. Note that we are in the setting of Lemma 5.3.2 with $A_k = |\varphi_k(Q_k) - \varphi_k(1)|$, $B_k = F_k^{-1}$ and $C_k = \mu_k^{-1}$ and with the sequence $A_k, k \geq 1$ being identically distributed. As seen in the proof of Lemma 5.4.3 we have $\beta_N + \gamma_N = \varepsilon_N + o(\varepsilon_N)$. Therefore, in view of Lemma 5.3.2 and Remark 5.3.3 it is sufficient to show that $\mathbb{E}_N [|\varphi_1(Q_1) - \varphi_1(1)|] \rightarrow 0$. Let $S = 1 - \eta^2 F_1'(1)^3$ and $D = \frac{1}{\eta F_1'(1)}$ with $\eta > 0$. On the event $\{0 < S < Q_1\}$ we have by Lemma 2 in [Ker20]

$$\begin{aligned} |\varphi_1(Q_1) - \varphi_1(1)| &\leq 2F_1'(1) \frac{F_1''(1)^2}{F_1'(1)^4} (1 - S) + 2D \frac{F_1''(1)}{F_1'(1)^2} (1 - S) \\ &\quad + \frac{2}{F_1'(1)^2} \mathbb{E}_N [\xi^2; \xi > D \mid V] \\ &\leq 2\eta^2 F_1''(1)^2 + 2\eta F_1''(1) + 2\eta^2 \mathbb{E}_N [\xi^4 \mid V]. \end{aligned} \quad (5.22)$$

Taking the expectation in (5.22) one obtains by Lemma 1 in [Ker20]

$$\mathbb{E}_N [|\varphi_1(Q_1) - \varphi_1(1)|] \leq 2\eta \mathbb{E}_N [\eta F_1''(1)^2 + F_1''(1) + \eta \xi^4] + 2\mathbb{E}_N [\varphi_1(1); \{0 < S < Q_1\}^c].$$

Since η can be chosen arbitrarily small and (5.1), it remains to show $\mathbb{E}_N [\varphi_1(1); \{0 < S < Q_1\}^c] \rightarrow 0$. Because of (5.18) $\varphi_1(1)$ is uniformly integrable, thus it suffices to show $\mathbb{P}_N (\{0 < S < Q_1\}^c) \rightarrow 0$. Note that $S \rightarrow 1 - \eta^2$ in probability in view of (5.2), thus to finish the proof we show $Q_1 \rightarrow 1$ in probability.

As a consequence of Lemma 5.4.4 it follows $\mathbb{E}_N \left[\frac{1}{X} \right] \rightarrow 0$. Then Proposition 5.2.1 implies that $\pi_N \rightarrow 0$, hence $Q_1 \rightarrow 1$ in probability, provided that the assumptions of Proposition 5.2.1 are satisfied.

From the proof of Lemma 5.3.2, it follows that for $u > 0$ sufficiently small,

$$\mathbb{E}_N [B^u] = \mathbb{E}_N [F'(1)^{-u}] < 1.$$

Therefore, via Jensen's inequality

$$\mathbb{E}_N [\log F'(1)] = -\frac{1}{u} \mathbb{E}_N [\log F'(1)^{-u}] > -\frac{1}{u} \log \mathbb{E}_N [F'(1)^{-u}] > 0.$$

The remaining assumptions of Proposition 5.2.1 are immediate from (5.1) and (5.2). \square

Proof of Theorem 10. i): Here, $\frac{\nu_N}{\varepsilon_N} \rightarrow 0$. From Lemma 5.4.3 and 5.4.5 we get

$$\frac{1}{\varepsilon_N X} \rightarrow \frac{2}{\sigma^2} \quad \text{in probability.}$$

Since $\frac{1}{\varepsilon_N X}$ is uniformly integrable (Lemma 5.4.4) it follows

$$\mathbb{E}_N \left[\frac{1}{\varepsilon_N X} \right] \rightarrow \frac{2}{\sigma^2}.$$

By Proposition 5.2.1 the proof is concluded.

ii): Now we have $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho \in (0, 2)$. By the same line of arguments we obtain

$$\mathbb{E}_N \left[\frac{1}{\nu_N X} \right] \rightarrow \frac{b^a}{\Gamma(a)} \int_0^\infty \frac{1}{x} x^{-a-1} e^{-\frac{b}{x}} dx = \frac{a}{b} = \frac{2(1-\rho) + \rho}{\rho\sigma^2}.$$

By Proposition 5.2.1 we get that

$$\mathbb{P}_N (Z_\infty > 0) = \nu_N \mathbb{E}_N \left[\frac{1}{\nu_N X} \right] \sim \nu_N \frac{2}{\sigma^2} \frac{1-\frac{\rho}{2}}{\rho} \sim \varepsilon_N \frac{2}{\sigma^2} \left(1 - \frac{\rho}{2}\right),$$

which gives the claim.

iii): We want to confirm that $\mathbb{E}_N [\log F'(1)] < 0$ holds for large N , which implies that the branching process is subcritical and dies out almost surely, see [SW69]. To this end we prove

$$\mathbb{E}_N [\log F'(1)] = \varepsilon_N - \frac{1}{2} \nu_N + o(\varepsilon_N), \quad (5.23)$$

which indeed implies $\mathbb{E}_N [\log F'(1)] < 0$ for N large enough, since we have $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$, with $\rho > 2$.

We split the expectation $\mathbb{E}_N [\log F'(1)]$ into two parts and obtain for any $\eta > 0$

$$\begin{aligned} \mathbb{E}_N [\log F'(1); |F'(1) - 1| > \eta] &\leq \mathbb{E}_N [(\log F'(1))^2]^{\frac{1}{2}} \mathbb{P}_N (|F'(1) - 1| > \eta)^{\frac{1}{2}} \\ &\leq \mathbb{E}_N \left[\left(\frac{1}{F'(1)} + F'(1) \right)^2 \right]^{\frac{1}{2}} \frac{1}{\eta^{2+\delta/2}} \mathbb{E}_N [|F'(1) - 1|^{4+\delta}]^{\frac{1}{2}} \\ &= O(\varepsilon_N^{2+\delta/2} + \nu_N^{1+\delta/2}) = o(\varepsilon_N), \end{aligned}$$

because of (5.2). By a Taylor expansion we have for some U between 1 and $F'(1)$

$$\log F'(1) = (F'(1) - 1) - \frac{1}{2}(F'(1) - 1)^2 U^{-2}.$$

Therefore,

$$\begin{aligned} & \left| \mathbb{E}_N [\log F'(1); |F'(1) - 1| \leq \eta] - \mathbb{E}_N \left[(F'(1) - 1) - \frac{1}{2}(F'(1) - 1)^2 \right] \right| \\ & \leq \left| \mathbb{E}_N \left[\frac{1}{2}(F'(1) - 1)^2 (U^{-2} - 1); |F'(1) - 1| \leq \eta \right] \right| \\ & \quad + \left| \mathbb{E}_N \left[(F'(1) - 1) - \frac{1}{2}(F'(1) - 1)^2; |F'(1) - 1| > \eta \right] \right| \\ & \leq \mathbb{E}_N \left[\frac{1}{2}(F'(1) - 1)^2 ((1 - \eta)^{-2} - 1) \right] + O(\mathbb{E}_N [|F'(1) - 1|^{2+\delta}]) \\ & = \frac{1}{2}(\nu_N + \varepsilon_N^2)((1 - \eta)^{-2} - 1) + o(\varepsilon_N) \end{aligned}$$

Altogether this implies

$$\left| \mathbb{E}_N [\log F'(1)] - \mathbb{E}_N \left[(F'(1) - 1) - \frac{1}{2}(F'(1) - 1)^2 \right] \right| \leq \frac{1}{2}\nu_N((1 - \eta)^{-2} - 1) + o(\varepsilon_N),$$

Letting $\eta \rightarrow 0$ this proves (5.23). □

Bibliography

- [Als21] G. Alsmeyer. Linear fractional Galton-Watson processes in random environment and perpetuities. *Preprint available at <https://arxiv.org/abs/2109.15007>*, 2021.
- [AIR09] G. Alsmeyer, A. Iksanov, and U. Rösler. On distributional properties of perpetuities. *J. Theor. Probab.*, 22(3):666–682, 2009.
- [ABSS16] E. Anceaume, Y. Busnel, E. Schulte-Geers, and B. Sericola. Optimization results for a generalized coupon collector problem. *J. Appl. Probab.*, 53(2):622–629, 2016.
- [Ath92] K. B. Athreya. Rates of decay for the survival probability of a mutant gene. *J. Math. Biol.*, 30(6):577–581, 1992.
- [AK71a] K. B. Athreya and S. Karlin. On branching processes with random environments. I. Extinction probabilities. *Ann. Math. Statist.*, 42:1499–1520, 1971.
- [AK71b] K. B. Athreya and S. Karlin. Branching processes with random environments. II. Limit theorems. *Ann. Math. Statist.*, 42:1843–1858, 1971.
- [AN72] K. B. Athreya and P. E. Ney. *Branching processes*. Springer-Verlag, New York-Heidelberg, 1972. Grundlehren Math. Wiss., Band 196.
- [AS05] S. R. Athreya and J. M. Swart. Branching-coalescing particle systems. *Probab. Theory Related Fields*, 131(3):376–414, 2005.
- [BB20] E. Baake and M. Baake. Ancestral lines under recombination. *in: Probabilistic Structures in Evolution*, E. Baake, A. Wakolbinger, eds., EMS Press Berson, 2021, pp 365-382., 2020.
- [BCH18] E. Baake, F. Cordero, and S. Hummel. A probabilistic view on the deterministic mutation-selection equation: dynamics, equilibria, and ancestry via individual lines of descent. *J. Math. Biol.*, 77(3):795–820, 2018.
- [BGPW19] E. Baake, A. González Casanova, S. Probst, and A. Wakolbinger. Modelling and simulating Lenski’s long-term evolution experiment. *Theor. Popul. Biol.*, 127:58–74, 2019.
- [BLW16] E. Baake, U. Lenz, and A. Wakolbinger. The common ancestor type distribution of a Λ -Wright-Fisher process with selection and mutation. *Electron. Commun. Probab.*, 21:59, 16, 2016.
- [BW21] E. Baake and A. Wakolbinger, editors. *Probabilistic Structures in Evolution*. EMS Press, 2021.

- [BD95] F. Ball and P. Donnelly. Strong approximations for epidemic models. *Stochastic Process. Appl.*, 55(1):1–21, 1995.
- [BB09] V. Bansaye and J. Berestycki. Large deviations for branching processes in random environment. *Markov Process. Related Fields*, 15(4):493–524, 2009.
- [BB14] V. Bansaye and C. Böinghoff. Small positive values for supercritical branching processes in random environment. *Ann. Inst. Henri Poincaré Probab.*, 50(3):770–805, 2014.
- [BR13] A. D. Barbour and G. Reinert. Approximating the epidemic curve. *Electron. J. Probab.*, 18:1–30, 2013.
- [Ber09] N. Berestycki. *Recent progress in coalescent theory*, volume 16 of *Ensaio Matemáticos*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009.
- [Bie45] I. J. Bienaymé. De la loi de multiplication et de la durée des familles. *Soc. Philomat. Paris Extraits, Sér.*, 5(37-39):4, 1845.
- [BBE13] M. Birkner, J. Blath, and B. Eldon. An ancestral recombination graph for diploid populations with skewed offspring distribution. *Genetics*, 193(1):255–290, 2013.
- [BLS18] M. Birkner, H. Liu, and A. Sturm. Coalescent results for diploid exchangeable population models. *Electron. J. Probab.*, 23(49):44, 2018.
- [BGPW21a] F. Boenkost, A. González Casanova, C. Pokalyuk, and A. Wakolbinger. Haldane’s formula in Cannings models: The case of moderately weak selection. *Electron. J. Probab.*, 26(4):36, 2021.
- [BGPW21b] F. Boenkost, A. González Casanova, C. Pokalyuk, and A. Wakolbinger. Haldane’s formula in Cannings models: The case of moderately strong selection. *J. Math. Biol.*, 83(70):31, 2021.
- [BK21] F. Boenkost and G. Kersting. Haldane’s asymptotics for supercritical branching processes in an iid random environment. Submitted, preprint available at <https://arxiv.org/abs/2109.10684>, 2021.
- [BS98] E. Bolthausen and A. S. Sznitman. On Ruelle’s probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.
- [BDM16] D. Buraczewski, E. Damek, and T. Mikosch. *Stochastic models with power-law tails. The equation $X = AX + B$* . Springer Series in Operations Research and Financial Engineering. Springer, 2016.
- [Can74] C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach, I. haploid models. *Adv. in Appl. Prob.*, 6(2):260–290, 1974.
- [Can75] C. Cannings. The latent roots of certain Markov chains arising in genetics: a new approach, II. further haploid models. *Adv. in Appl. Prob.*, 7(2):264–282, 1975.
- [Cor17] F. Cordero. Common ancestor type distribution: A Moran model and its deterministic limit. *Stoch. Processes Appl.*, 127.2, 2017.

- [CM19] F. Cordero and M. Möhle. On the stationary distribution of the block counting process for population models with mutation and selection. *J. Math. Anal. Appl.*, 474(2):1049–1081, 2019.
- [Cra38] H Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. *Actual. Sci. Ind.* 736: 5-23., 1938.
- [Dar59] C. Darwin. *On the origin of species*. 1859.
- [DZ94] A Dembo and O Zeitouni. Large deviations techniques and applications. *Springer Verlag New York*, 2nd edition, 1994.
- [DF07] M. M. Desai and D. S. Fisher. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798, 2007.
- [DK99] P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27(1):166–205, 1999.
- [Dur08] R. Durrett. *Probability models for DNA sequence evolution*. Springer Science & Business Media, 2008.
- [DM11] R. Durrett and J. Mayberry. Traveling waves of selective sweeps. *Ann. Appl. Probab.*, 21(2):699–744, 2011.
- [Esh81] I. Eshel. On the survival probability of a slightly advantageous mutant gene with a general distribution of progeny size—a branching process model. *J. Math. Biol.*, 12(3):355–362, 1981.
- [Esh84] I. Eshel. On the survival probability of a slightly advantageous mutant gene in a multitype population: a multidimensional branching process model. *J. Math. Biol.*, 19(2):201–209, 1984.
- [Eth11] A. Etheridge. *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011.
- [EPW09] A. M. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. How often does the ratchet click? Facts, heuristics, asymptotics. In *Trends in stochastic analysis*, volume 353 of *London Math. Soc. Lecture Note Ser.*, pages 365–390. Cambridge Univ. Press, Cambridge, 2009.
- [EK86] S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, Inc., 1986.
- [Ewe04] W.J. Ewens. *Mathematical Population Genetics: I. Theoretical Introduction, Second Edition*. Springer Science + Business Media New York, 2004.
- [Fel71] W. Feller. *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons, second edition, 1971.
- [Fis23] R. A. Fisher. On the dominance ratio. *Proc. R. Soc. Edin.* 42, 1923.
- [Fis30] R. A. Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, 1930.
- [Gal73] F. Galton. Problem 4001. *Educational Times*, 1:17, April 1873.

- [GW74] F. Galton and H. W. Watson. On the probability of the extinction of families. *J. Roy. Anthropol. Inst.*, 4:138–144, 1874.
- [GK00] J. Geiger and G. Kersting. The survival probability of a critical branching process in random environment. *Theory Probab. Appl.*, 45(3):607–615, 2000.
- [Gil74] J. H. Gillespie. Natural selection for within-generation variance in offspring number. *Genetics*, 76(3):601–606, 1974.
- [Gil75] J. H. Gillespie. Natural selection for within-generation variance in offspring number II. discrete haploid models. *Genetics*, 81(2):403–413, 1975.
- [GZ08] P. W. Glynn and A. Zeevi. Bounding stationary expectations of Markov processes. In *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, volume 4 of *Inst. Math. Stat. (IMS) Collect.*, pages 195–214. Inst. Math. Statist., Beachwood, OH, 2008.
- [GKWY17] A. González Casanova, N. Kurt, A. Wakolbinger, and L. Yuan. An individual-based model for the Lenski experiment, and the deceleration of the relative fitness. *Stoch. Processes Appl.*, 126:2211–2252, 2017.
- [GS18] A. González Casanova and D. Spanò. Duality and fixation in Ξ -Wright-Fisher processes with frequency-dependent selection. *Ann. Appl. Probab.*, 28:250–284, 2018.
- [GLM17] I. Grama, Q. Liu, and E. Miqueu. Berry-Esseen’s bound and Cramér’s large deviation expansion for a supercritical branching process in a random environment. *Stochastic Process. Appl.*, 127(4):1255–1281, 2017.
- [GM97] R. C. Griffiths and P. Marjoram. An ancestral recombination graph. In *Progress in population genetics and human evolution (Minneapolis, MN, 1994)*, volume 87 of *IMA Vol. Math. Appl.*, pages 257–270. Springer, New York, 1997.
- [HJV05] P. Haccou, P. Jagers, and V. A. Vatutin. *Branching processes: variation, growth, and extinction of populations*. Number 5. Cambridge university press, 2005.
- [Hal27] J. B. S. Haldane. A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Proc. Camb. Philos. Soc.*, 23(7):838–844, 1927.
- [Ham92] B. Hambly. On the limiting distribution of a supercritical branching process in a random environment. *J. Appl. Probab.*, 29(3):499–518, 1992.
- [Har63] T. E. Harris. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer, 1963.
- [Hoe94] W. Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.
- [Hop92] F. M. Hoppe. Asymptotic rates of growth of the extinction probability of a mutant gene. *J. Math. Biol.*, 30(6):547–566, 1992.

- [HL14] C. Huang and Q. Liu. Convergence in L^p and its exponential rate for a branching process in a random environment. *Electron. J. Probab.*, 19:no. 104, 22, 2014.
- [HM21] T. Huillet and M. Möhle. Asymptotic genealogies for a class of generalized Wright–Fisher models. Preprint available at <https://arxiv.org/abs/2106.10939v1>, 2021.
- [Jan18] S. Janson. Tail bounds for sums of geometric and exponential variables. *Stat Probabil Lett*, 135:1–6, 2018.
- [JDP83] K. Joag-Dev and F. Proschan. Negative association of random variables, with applications. *Ann. Stat.*, 11(1):286–295, 1983.
- [Ken66] D. G. Kendall. Branching processes since 1873. *J. London Math. Soc.*, 41:385–406. (1 plate), 1966.
- [Ken75] D. G. Kendall. The genealogy of genealogy: branching processes before (and after) 1873. with a french appendix containing bienaymé’s paper of 1845. *Bull. London Math. Soc.*, 7(3):225–253, 1975.
- [Ker20] G. Kersting. A unifying approach to branching processes in varying environments. *J. Appl. Probab.*, 57:196–200, 2020.
- [KV17] G. Kersting and V. A. Vatutin. *Discrete time branching processes in random environment*. Wiley, 2017.
- [Kim57] M. Kimura. Some problems of stochastic processes in genetics. *Ann. Math. Statistics*, 28:882–901, 1957.
- [Kim62] M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713, 1962.
- [Kin82a] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13(3):235–248, 1982.
- [Kin82b] J. F. C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, (Special Vol. 19A):27–43, 1982.
- [KM10] A. Klenke and L. Mattner. Stochastic ordering of classical discrete distributions. *Adv. in Appl. Probab.*, 42(2):392–410, 2010.
- [Kol38] A.N. Kolmogorov. Zur Lösung einer biologischen Aufgabe. *Comm. Math. Mech. Chebyshev Univ. Tomsk*, 2(1):1–12, 1938.
- [KN97] S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theor. Popul. Biol.*, 51:210–237, 1997.
- [Kur70] T.G. Kurtz. Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.*, 7, 1970.
- [Kur81] T.G. Kurtz. Approximation of population processes. *CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM*, 36, 1981.
- [LRST91] R. E. Lenski, M. R. Rose, S. C. Simpson, and S. C. Tadler. Long-term experimental evolution in *Escherichia coli*. I. adaptation and divergence during 2,000 generations. *Am. Nat.*, 138(6):1315–1341, 1991.

- [LT94] R. E. Lenski and M. Travisano. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. USA*, 91(15):6808–6814, 1994.
- [LKBW15] U. Lenz, S. Kluth, E. Baake, and A. Wakolbinger. Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution. *Theor. Popul. Biol.*, 103:27–37, 2015.
- [LL07] S. Lessard and V. Ladret. The probability of fixation of a single mutant in an exchangeable selection model. *J. Math. Biol.*, 54(5):721–744, 2007.
- [LP17] R. Lyons and Y. Peres. *Probability on trees and networks*. Vol. 42. Cambridge University Press, 2017.
- [Man09] S. Mano. Duality, ancestral and diffusion processes in models with selection. *Theor. Popul. Biol.*, 75(2-3):164–175, 2009.
- [Men65] G. Mendel. Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, IV:3–47, 1865.
- [MN98] T. Mikosch and A. V. Nagaev. Large deviations of heavy-tailed sums with applications in insurance. *Extremes*, 1(1):81–110, 1998.
- [Möh00] M. Möhle. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.*, 32:983–993, 2000.
- [Mor58] P. A. P. Moran. Random processes in genetics. *Proc. Cambridge Philos. Soc.*, 54:60–71, 1958.
- [Mor59] P. A. P. Moran. The survival of a mutant gene under selection. *J. Austral. Math. Soc.*, 1(part 1):121–126, 1959.
- [Mor60] P. A. P. Moran. The survival of a mutant gene under selection. II. *J. Austral. Math. Soc.*, 1:485–491, 1960.
- [NK97] C. Neuhauser and S. M. Krone. The genealogy of samples in models with selection. *Genetics*, 145(2):519–534, 1997.
- [OH21] T. Okada and O. Hallatschek. Dynamic sampling bias and overdispersion induced by skewed offspring distributions. *Genetics*, *iyab135*, 2021.
- [PW08] Z. Patwa and L. Wahl. The fixation probability of beneficial mutations. *J. R. Soc. Interface*, 5(28):1279–1289, 2008.
- [Pit99] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902, 1999.
- [PP13] C. Pokalyuk and P. Pfaffelhuber. The ancestral selection graph under strong directional selection. *Theor. Popul. Biol.*, 87:25–33, 2013.
- [PW20] C. Pokalyuk and A. Wakolbinger. Maintenance of diversity in a hierarchical host-parasite model with balancing selection and reinfection. *Stochastic Process. Appl.*, 130(2):1119–1158, 2020.
- [Roo69] D. H. Root. The existence of certain stopping times on Brownian motion. *Ann. Math. Statist.*, 40:715–718, 1969.

- [RBW08] I. M. Rouzine, É. Brunet, and C. O. Wilke. The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. *Theor. Popul. Biol.*, 73(1):24–46, 2008.
- [Rue87] D. Ruelle. A mathematical reformulation of Derrida's REM and GREM. *Comm. Math. Phys.*, 108(2):225–239, 1987.
- [Sag99] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36(4):1116–1125, 1999.
- [Sch03] J. Schweinsberg. Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch. Processes Appl.*, 106(1):107–139, 2003.
- [Sla68] R. S. Slack. A branching process with mean one and possibly infinite variance. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 9:139–145, 1968.
- [Smi68] W. L. Smith. Necessary conditions for almost sure extinction of a branching process with random environment. *Ann. Math. Stat.*, 39(6):2136–2140, 1968.
- [SW69] W. L. Smith and W. E. Wilkinson. On branching processes in random environments. *Ann. Math. Statist.*, 40:814–827, 1969.
- [Tan77] D. Tanny. Limit theorems for branching processes in a random environment. *Ann. Probab.*, 5(1):100–116, 1977.
- [Tan88] D. Tanny. A necessary and sufficient condition for a branching process in a random environment to grow like the product of its means. *Stochastic Process. Appl.*, 28(1):123–139, 1988.
- [Tay07] J. E. Taylor. The common ancestor process for a Wright-Fisher diffusion. *Electron. J. Probab.*, 12:no. 28, 808–847, 2007.
- [VZ93] V. A. Vatutin and A. M. Zubkov. Branching processes. II. *J. Sov. Math*, 67(6):3407–3485, 1993.
- [Wri31] S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97, 1931.

Appendix A

Zusammenfassung

A.1 Haldanes Asymptotik für leicht superkritische Prozesse

Die Theorie der mathematischen Populationsgenetik ist die Synthese der von Darwin begründeten Evolutionstheorie [Dar59] und der Mendelschen Vererbungslehre [Men65] und wurde maßgeblich von Fisher, Haldane und Wright zu Beginn des zwanzigsten Jahrhunderts begründet. Ein Ziel der Populationsgenetik ist es die Evolution eines Typs (oder Allels) in einem idealisierten mathematischen Modell zu beschreiben, wobei die Population unterschiedlichen evolutionären Mechanismen wie Gendrift, Selektion, Mutation oder Migration ausgesetzt ist.

In der vorliegenden Arbeit wird vor allem der Zusammenhang zwischen genetischem Drift und Selektion betrachtet. Dies geschieht unter verschiedenen Annahmen an die Stärke dieser Einflüsse. Wir analysieren dabei wie sich die Fixationswahrscheinlichkeit in Abhängigkeit des selektiven Vorteils asymptotisch verhält.

Die klassische Frage nach der Fixationswahrscheinlichkeit (also die Wahrscheinlichkeit dass sich der bevorzugte Typ in der ganzen Population durchsetzt) geht bereits zurück auf Haldane um 1920 [Hal27], der die mittlerweile nach ihm benannte Asymptotik entwickelt hat. Betrachtet man einen bevorzugten Typ mit kleinem selektiven Vorteil s in einer großen Population, so ist die asymptotische Fixationswahrscheinlichkeit $\pi(s)$ gegeben durch

$$\pi(s) \sim \frac{2s}{\sigma^2}, \quad \text{für } s \downarrow 0$$

wobei σ^2 die Varianz der Nachkommenverteilung bezeichnet. Im Folgenden definieren wir zunächst kurz die grundlegenden stochastischen Modelle und stellen diese vor. Danach erläutern wir die Resultate aus [BGPW21a], [BGPW21b] und [BK21], die den Kapiteln 3 bis 5 entsprechen. Die Arbeiten sind in Zusammenarbeit mit Adrián González Casanova, Cornelia Pokalyuk und Anton Wakolbinger, sowie mit Götz Kersting entstanden. Abschließend geben wir noch einen kurzen Ausblick auf eine in Vorbereitung befindliche Arbeit, die auf einer Kollaboration mit Matthias Birkner, Iulia Dahmer und Cornelia Pokalyuk beruht.

A.1.1 Einleitung

In diesem Abschnitt werden wir den mathematischen Rahmen der theoretischen Populationsgenetik, der in dieser Arbeit benötigt wird, vorstellen und kommentieren. Unser Ziel ist es, die Überlebens- oder Fixationswahrscheinlichkeiten von individuen-basierten stochastischen Populationsmodellen zu untersuchen, die leicht superkritisch sind. Mit *leicht superkritisch* bezeichnen wir Prozesse, bei denen der bevorzugte Typ (im Gegensatz zum Wildtyp) in Erwartung $1 + s > 1$ Nachkommen hat

und wir betrachten den Grenzwert $s \downarrow 0$. Das wohl einfachste stochastische Modell für die Evolution einer Population ist der Galton-Watson-Prozess.

Definition A.1.1 (Galton-Watson-Prozess).

Wir nennen $(Z_n, n \geq 0)$ einen Galton-Watson-Prozess, falls Z_n die Rekursion

$$Z_n = \sum_{j=1}^{Z_{n-1}} \xi_j^{(n)}, \quad Z_0 = z_0 \geq 1$$

erfüllt, wobei $(\xi_j^{(n)}, j \geq 1, n \geq 1)$ unabhängige Kopien einer auf \mathbb{N}_0 verteilten Zufallsvariablen ξ sind.

Galton-Watson-Prozesse werden mithilfe des Erwartungswertes $m = \mathbb{E}[\xi]$ der Nachkommenverteilung klassifiziert, siehe [AN72] oder [Har63]. Falls der Galton-Watson-Prozess subkritisch ($m < 1$) oder kritisch ($m = 1$) ist, stirbt der Prozess fast sicher aus (abgesehen vom trivialen Fall $\mathbb{P}(\xi = 1) = 1$). Nur im superkritischen Fall ($m > 1$) ist die Überlebenswahrscheinlichkeit größer als 0. Sei

$$f(t) = \sum_{k=0}^{\infty} t^k \mathbb{P}(\xi = k), \quad t \in [0, 1]$$

die wahrscheinlichkeitserzeugende Funktion. Die Überlebenswahrscheinlichkeit $\pi = 1 - q$, kann durch die kleinste Lösung der Fixpunktgleichung

$$q = f(q) \tag{A.1}$$

ausgedrückt werden. Im Folgenden betrachten wir oft Folgen von Prozessen die leicht superkritisch sind. Für einen leicht superkritischen Galton-Watson-Prozess mit $m = 1 + s$ und endlicher Varianz σ^2 , ist die Überlebenswahrscheinlichkeit $\pi(s)$ durch Haldanes Asymptotik gegeben

$$\pi(s) \sim \frac{2s}{\sigma^2}, \quad \text{für } s \downarrow 0.$$

Diese Asymptotik wurde zum Beispiel unter der Annahme gleichmäßig beschränkter dritter Momente [Kol38], [Ath92] mithilfe der Charakterisierung der Überlebenswahrscheinlichkeit in (A.1) gezeigt, siehe auch [Esh81] und [Hop92]. Wir nennen $s > 0$ den *selektiven Vorteil* der bevorzugten Individuen. Im Abschnitt A.1.7 werden wir auf eine Erweiterung von Galton-Watson-Prozessen zurückkommen und die Überlebenswahrscheinlichkeit von Verzweigungsprozessen in zufälliger Umgebung betrachten. Die Verwendung von reinen Galton-Watson-Prozessen zur Modellierung der Evolution einer Population hat den Nachteil, dass die Population im Überlebensfall über alle Grenzen hinaus wächst, was in der Realität offensichtlich nicht beobachtet wird.

Ein einfaches Modell zur Beschreibung einer Population mit konstanter Populationsgröße N in diskreter Zeit, ist das *Cannings-Modell* [Can74] und [Can75]. Dazu definieren wir zunächst den Begriff der *Austauschbarkeit*.

Definition A.1.2 (Austauschbarkeit).

Wir nennen die Zufallsvariablen X_1, \dots, X_n austauschbar, falls für alle Permutationen $\psi : [n] \rightarrow [n]$ gilt.

$$(X_1, \dots, X_n) \stackrel{d}{=} (X_{\psi(1)}, \dots, X_{\psi(n)}).$$

Mit dieser Definition können wir die Klasse der Cannings-Modelle einführen. Im folgenden bezeichnen wir mit $\nu = (\nu_1, \dots, \nu_N)$ den zufälligen Kinderzahlvektor, also den Vektor der die Anzahl Nachkommen je Individuum beschreibt.

Definition A.1.3 (Cannings-Modell).

Sei N die feste Populationsgröße. Bezeichne mit $\nu^{(g)}$ den Kinderzahlvektor in Generation g . Wir nehmen an, dass die Verteilung $\mathcal{L}(\nu^{(g)})$ nicht von der Generation g abhängt und dass die $\nu^{(g)}$, $g \in \mathbb{Z}$ unabhängig sind. Weiterhin sei $\nu^{(g)}$ austauschbar, mit $\sum_{i=1}^N \nu_i^{(g)} = N$ für alle $g \in \mathbb{Z}$. Das resultierende Populationsmodell nennen wir Canning-Modell.

Der Spezialfall, dass ν multinomialverteilt mit Parametern $(N; \frac{1}{N}, \dots, \frac{1}{N})$ ist, ist als das *Wright-Fisher-Modell* [Fis30] und [Wri31] bekannt. Die Klasse der neutralen Cannings-Modelle mit einem selektiven Mechanismus auszustatten ist nicht wie bei anderen Modellen direkt möglich. Es ist nicht offensichtlich wie die Verteilung von ν so angepasst werden kann, dass die bevorzugten Individuen einen selektiven Vorteil haben, vergleiche hierzu den Ansatz von Lessard und Ladret [LL07]. Beim Wright-Fisher Modell geschieht dies, indem man den bevorzugten Individuen das Gewicht $1 + s$ zuweist, wohingegen die Wildtyp Individuen das Gewicht 1 erhalten. Der Kinderzahlvektor entsteht dann wieder durch multinomiales Ziehen, wobei die Wahrscheinlichkeit ein Individuum als Elter zu ziehen proportional zu seinem Gewicht ist.

Im Abschnitt A.1.2 werden wir eine Möglichkeit aufzeigen selektive Cannings-Modelle zu definieren, die den eben beschriebenen Ansatz für Wright-Fisher-Modelle mit Selektion aufgreift. Dies wird auf eine Art und Weise getan, die es erlaubt den *Cannings-Anzahlprozess* (die Anzahl der Wildtyp-Individuen in jeder Generation) vorwärts in der Zeit zu beschreiben als auch einen dualen Rückwärts-Prozess nach Ideen von Krone und Neuhauser [KN97], [NK97] und González Casanova und Spanò [GS18] zu definieren.

Alle bisher in diesem Abschnitt beschriebenen Populationsmodelle modellieren die zeitliche Entwicklung einer Population vorwärts in der Zeit. Das bedeutet, wenn die Verteilung der Typen in der vorherigen Generation bekannt ist, so kann die Verteilung der Typen in der nächsten Generation bestimmt werden. Im Kontext der mathematischen Populationsgenetik ist es ein weit verbreiteter Ansatz, in der Zeit zurückzublicken und die Ahnenlinien einer Stichprobe von Individuen zu verfolgen.

Angenommen, wir befinden uns in einem neutralen Wright-Fisher-Modell mit zwei Typen und wir möchten den Typ einer Stichprobe von n Individuen bestimmen. Wir können das Modell entweder zeitlich vorwärts betrachten und die Typen der n Individuen beobachten oder wir können die Ahnenlinien der n Individuen zeitlich rückwärts verfolgen und die Typen der Vorfahren bestimmen, was uns folglich die Typenverteilung der Stichprobe liefert. Immer wenn zwei Individuen denselben Elter wählen, verschmelzen ihre Ahnenlinien, daher sprechen wir von *Koaleszenten*.

Einer der bekanntesten Koaleszenten wurde um 1980 von Kingman eingeführt [Kin82a]. Der *Kingman-Koaleszent* beschreibt (unter anderem) den Grenzwert der reskalierten Genealogie einer Stichprobe im neutralen Wright-Fisher-Modells, nachdem die Zeit passend beschleunigt wurde. Dazu sei \mathcal{P}_n die Menge aller Partitionen von $[n]$.

Definition A.1.4 (Kingman-Koaleszent).

Wir nennen einen \mathcal{P}_n -wertigen Markov-Prozess $(\Pi_n(t))_{t \geq 0}$ *Kingmans-Koaleszent*, falls $\Pi_n(0)$ die Partition in n Blöcke $\{\{1\}, \dots, \{n\}\}$ ist und jedes Paar von Blöcken unabhängig voneinander mit Rate 1 verschmilzt.

Für neutrale Cannings-Modelle beschreibt der Kingman-Koaleszent die Genealogie im Grenzwert, falls Möhles Bedingung [Möh00] erfüllt ist, nämlich dass die Paarverschmelzungswahrscheinlichkeit von kleinerer Ordnung ist als die Wahrscheinlichkeit für Tripel-Verschmelzungen, also

$$\frac{\mathbb{E}[\nu_1(\nu_1 - 1)]}{N - 1} = o\left(\frac{\mathbb{E}[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{(N - 1)(N - 2)}\right).$$

Falls diese Bedingung nicht erfüllt ist, sind andere Genealogien möglich, wie zum Beispiel die Klasse der Λ -Koaleszenten welche unabhängig von einander von Pitman [Pit99], Sagitov [Sag99] und Donnelly und Kurtz [DK99] eingeführt wurden. Cannings-Modelle mit Selektion bei denen die neutrale reskalierte Genealogie gegen einen Λ -Koaleszenten konvergiert werden im Abschnitt A.1.6 betrachtet.

A.1.2 Cannings-Modelle mit moderater Selektion

Betrachten wir ein klassisches Cannings-Modell [Can74], [Can75] mit fester Populationsgröße $N \in \mathbb{N}$, das durch die Verteilung seiner austauschbaren und nicht negativen Nachkommenverteilung $\nu = (\nu_1, \dots, \nu_N)$, mit $\nu_1 + \dots + \nu_N = N$ definiert ist, siehe Definition A.1.3. Im Folgenden konzentrieren wir uns auf den Spezialfall von Cannings-Modellen, die eine *Paintbox-Konstruktion* ermöglichen.

Sei $\mathscr{W}^{(N)} = (W_1^{(N)}, \dots, W_N^{(N)})$ ein Vektor von zufälligen Gewichten, so dass $\mathscr{W}^{(N)}$ austauschbar ist und

$$\sum_{j=1}^N W_j^{(N)} = 1, \quad W_i^{(N)} \geq 0, i \in [N]$$

gilt. Um die Notation im Folgenden zu vereinfachen unterdrücken wir an manchen Stellen die Abhängigkeit von N . Manchmal schreiben wir auch $\mathscr{W}^{(g)}$ für den Vektor der zufälligen Gewichte in Generation g . Wir nehmen an, dass $(\mathscr{W}^{(g)})_{g \in \mathbb{Z}}$ eine Familie unabhängig und identisch verteilter Zufallsvariablen ist.

Die Anzahl der Nachkommen von einer Generation zur nächsten wird durch ein zweistufiges Verfahren erzeugt. Zunächst generieren wir den Vektor der zufälligen Gewichte \mathscr{W} . Gegeben \mathscr{W} sei dann ν multinomialverteilt mit Parametern $(N; \mathscr{W})$. Als Beispiel erhalten wir das klassische Wright-Fisher-Modell, wenn wir $W_i = \frac{1}{N}$ wählen, für alle $i \in [N]$.

Die genealogischen Beziehungen zwischen den Individuen der Generation $g - 1$ und g werden folgendermaßen beschrieben. Mit $(i, g), i \in [N]$ bezeichnen wir die Individuen der Generation g . Jedem einzelnen (i, g) wird ein Elter $(V(i, g), g - 1)$ in der Generation $g - 1$ zugewiesen, so dass für alle $j \in [N]$ gilt

$$\mathbb{P}\left(V(i, g) = j | \mathscr{W}^{(g-1)}\right) = W_j^{(g-1)}, \quad (\text{A.2})$$

gegeben $\mathscr{W}^{(g-1)}$ unabhängig für alle $i \in [N]$ und unabhängig zwischen den Generationen. Ein Vorteil dieser Definition ist die Existenz einer graphischen Darstellung für diese Klasse der Cannings-Modelle.

Für jedes $g \in \mathbb{Z}$ sei $U_1^{(g)}, \dots, U_N^{(g)}$ eine Folge unabhängig und identisch auf dem Einheitsquadrat $[0, 1] \times [0, 1]$ verteilter Zufallsvariablen. Wir teilen das Quadrat $[0, 1] \times [0, 1]$ in N Streifen $[0, 1] \times \mathcal{I}_1^{(g-1)}, \dots, [0, 1] \times \mathcal{I}_N^{(g-1)}$, so dass $\mathcal{I}_j^{(g-1)}$ die Länge

$W_j^{(g-1)}$ hat, also

$$\mathcal{I}_j^{(g-1)} = \left[\sum_{i=1}^{j-1} W_i^{(g-1)}, \sum_{i=1}^j W_i^{(g-1)} \right], \quad j \in [N].$$

Im Sinne von (A.2), erhalten wir die Gleichheit der Ereignisse

$$\{V(i, g) = j\} = \{U_i^{(g)} \in [0, 1] \times \mathcal{I}_j^{(g-1)}\},$$

dies ist illustriert in Abbildung A.1.

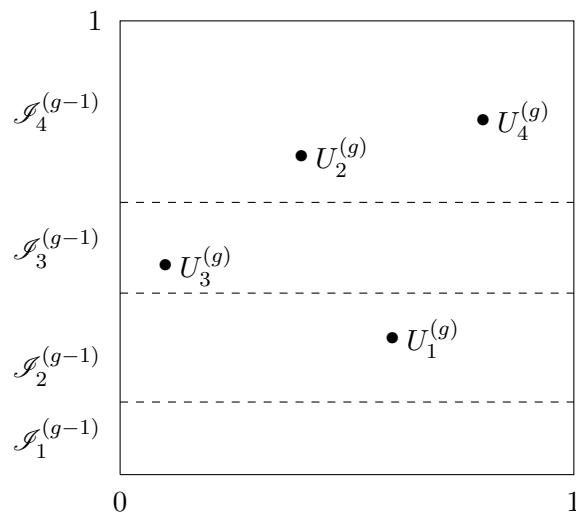


ABBILDUNG A.1: Eine Realisierung eines neutralen Cannings-Modells mit einer Paintbox-Darstellung und $N = 4$ Individuen. Zum Beispiel hat das Individuum 2 in der Generation g das Individuum 4 in der Generation $g - 1$ als Elternteil gewählt, da $U_2^{(g)} \in [0, 1] \times \mathcal{I}_4^{(g-1)}$.

Wir führen Selektion mit Stärke $s_N \in (0, 1)$ in diesem Modell folgendermaßen ein. Angenommen jedes Individuum ist entweder vom *Wildtyp* oder vom *bevorzugten Typ* und jedes Kind erbt den Typ seines Elters. Definiere den Vektor der modifizierten Gewichte $\widetilde{\mathcal{W}}^{(g-1)} = (\widetilde{W}_1^{(g-1)}, \dots, \widetilde{W}_N^{(g-1)})$ mit $\widetilde{W}_i^{(g-1)} = (1 - s_N)W_i^{(g-1)}$ wenn i vom Wildtyp ist und $\widetilde{W}_i^{(g-1)} = W_i^{(g-1)}$, falls i vom bevorzugten Typ ist, für alle $i \in [N]$. Auf die gleiche Weise wie zuvor ergibt sich die Generation g durch multinomiales Ziehen mit Gewichten proportional zu $\widetilde{\mathcal{W}}^{(g-1)}$. Also ist $(j, g - 1)$ der Elter des Individuums (i, g) mit Wahrscheinlichkeit

$$\mathbb{P}\left(V(i, g) = j \mid \widetilde{\mathcal{W}}^{(g-1)}\right) = \frac{\widetilde{\mathcal{W}}_j^{(g-1)}}{\sum_{k=1}^N \widetilde{\mathcal{W}}_k^{(g-1)}}, \quad j \in [N],$$

und gegeben $\widetilde{\mathcal{W}}^{(g-1)}$ unabhängig für alle $i \in [N]$ und unabhängig zwischen den Generationen.

Dieses Verfahren binden wir in die graphische Darstellung ein und definieren dafür die Mengen

$$\begin{aligned}\mathcal{B}^{(g-1)} &:= \{i \in [N] : (i, g-1) \text{ ist vom bevorzugten Typ}\}, \\ \mathcal{C}^{(g-1)} &:= \{i \in [N] : (i, g-1) \text{ ist vom Wildtyp}\}.\end{aligned}$$

Außerdem seien $(U_1^{(g)}, \dots, U_N^{(g)})$ unabhängig und uniform verteilt auf $\Gamma^{(g-1)}$, mit

$$\Gamma^{(g-1)} := \bigcup_{j \in \mathcal{B}^{(g-1)}} [0, 1] \times \mathcal{I}_j^{(g-1)} \cup \bigcup_{j \in \mathcal{C}^{(g-1)}} [0, 1 - s_N] \times \mathcal{I}_j^{(g-1)}.$$

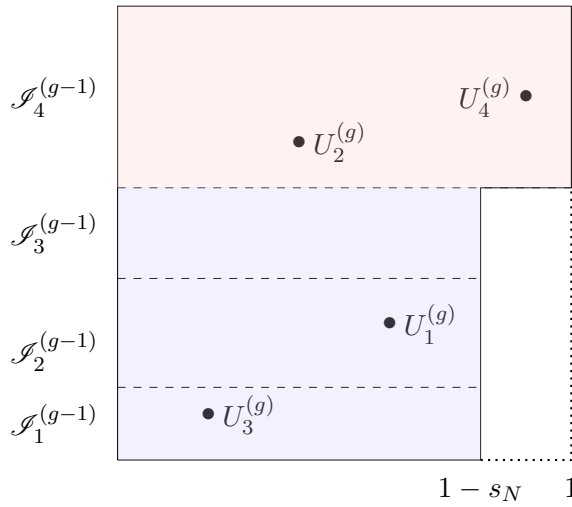


ABBILDUNG A.2: Beispiel einer Realisierung eines Übergangsschrittes eines Cannings-Modells mit $N = 4$ Individuen. Das Individuum 4 ist vom bevorzugten Typ, wohingegen die anderen Individuen vom Wildtyp sind.

Dies liefert die Gleichheit der Ereignisse (vergleiche mit dem Ausdruck in (A.2) und Abbildung A.2),

$$\begin{aligned}\{V(i, g) = j\} &= \{U_i^{(g-1)} \in [0, 1] \times \mathcal{I}_j^{(g-1)}\}, & \text{für } j \in \mathcal{B}^{(g-1)}, \\ \{V(i, g) = j\} &= \{U_i^{(g-1)} \in [0, 1 - s_N] \times \mathcal{I}_j^{(g-1)}\}, & \text{für } j \in \mathcal{C}^{(g-1)}.\end{aligned}$$

Im Folgenden betrachten wir das Aussterben der Wildtyp-Individuen (beziehungsweise das Überleben der selektiv bevorzugten Individuen). Um ihre Häufigkeit über die Generationen hinweg zu verfolgen, definieren wir den Zählprozess der Wildtyp-Individuen.

Definition A.1.5 (Cannings-Anzahl-Prozess).

Mit $\mathcal{K} = (K_g)_{g \geq 0}$ bezeichnen wir den Cannings-Anzahl-Prozess mit Parametern N, s_N und $\mathcal{L}(\mathcal{W})$ und Zustandsraum $[N]$. Dabei zählt K_g die Anzahl der Wildtyp Individuen in Generation g .

Aufgrund der Austauschbarkeit von \mathcal{W} hängen die Übergangswahrscheinlichkeiten von \mathcal{K} nicht von der genauen Konfiguration der bevorzugten Individuen in der vorherigen Generation ab, weswegen \mathcal{K} ein Markov-Prozess ist. Aufgrund

der zuvor beschriebenen Prozedur zur Generierung des Kinderzahlvektors sind die Übergangswahrscheinlichkeiten für K_g , gegeben $K_{g-1} = k$ und $\mathscr{W}^{(g-1)}$ die gleichen wie bei einer gemischt binomial verteilten Zufallsvariable mit Parametern N und $P(k, \mathscr{W})$, mit

$$P(k, \mathscr{W}) := \frac{(1 - s_N) \sum_{i=1}^k W_i}{(1 - s_N) \sum_{i=1}^k W_i + \sum_{i=k+1}^N W_i}.$$

A.1.3 Der ancestrale Selektionsprozess im Cannings-Modell

Für Cannings-Modelle mit Paintbox-Darstellung wie in Abschnitt A.1.2, existiert ein ancestraler Prozess, der dual zu \mathcal{K} ist. Der *Cannings ancestrale Selektionsprozess*, $\mathcal{A} = (A_m)_{m \in \mathbb{N}_0}$ mit den Parametern N, s_N und $\mathcal{L}(\mathscr{W})$, zählt die Anzahl potenzieller Eltern von einer Stichprobe von Individuen m Generationen in der Vergangenheit.

Der Cannings ancestrale Selektionsprozess ist eine $[N]$ -wertige Markovkette, deren Übergangswahrscheinlichkeiten durch einen Verzweigungs- und einem Verschmelzungsschritt beschrieben werden. Gegeben $A_m = a$, verzweigt der Prozess zunächst in die zufällige Summe $H = \sum_{i=1}^a G^{(i)}$, wobei $G^{(i)}$ unabhängige geometrisch verteilte Zufallsvariablen mit Parameter $1 - s_N$ sind. Also ist H negativbinomialverteilt mit Parametern a und $1 - s_N$. Für den Verschmelzungsschritt werden H Kugeln auf N Boxen je mit Wahrscheinlichkeit $\mathscr{W} = (W_1, \dots, W_N)$ verteilt. Die Verteilung von A_{m+1} gegeben $A_m = a$ ist die gleiche wie die Verteilung der Anzahl belegter Boxen nach diesem zweistufigen Verfahren.

Dieses Verfahren entsteht aufgrund der folgenden Beobachtungen auf natürliche Weise. Da der Prozess in der Zeit zurückblickt, gibt es keine Informationen darüber, wie viele bevorzugte Individuen in der vorherigen Generation vorhanden sind. Deswegen wird jedes Gewicht W_i in einen neutralen Anteil $(1 - s_N)W_i$ und einen selektiven Anteil $s_N W_i$ aufgeteilt. Dies führt dazu, dass das Quadrat $[0, 1] \times [0, 1]$ vertikal in einen neutralen Teil \mathcal{N} und einen selektiven Teil \mathcal{S} aufgeteilt wird,

$$\mathcal{N} := [0, 1 - s_N] \times [0, 1] \quad \mathcal{S} := [1 - s_N, 1] \times [0, 1].$$

Für jedes Individuum $j \in [N]$ in der Generation g betrachten wir nun eine Folge uniform auf dem Einheitsquadrat verteilter Zufallsvariablen $U_j^{(1,g)}, U_j^{(2,g)}, \dots$ die unabhängig und identisch verteilt sind. Für jedes Individuum warten wir bis die erste uniforme Zufallsvariablen in den Bereich \mathcal{N} fällt und bezeichnen den entsprechenden Index mit $\gamma(j, g)$,

$$\gamma(j, g) := \min\{\ell \geq 1 : U_j^{(\ell,g)} \in \mathcal{N}\}.$$

Also ist $\gamma(j, g)$ geometrisch verteilt mit Parameter $1 - s_N$ und wir nennen die Individuen $k \in [N]$ mit

$$U_j^{(\ell,g)} \in [0, 1] \times \mathcal{S}_k^{(g-1)}, \quad \text{für ein } 1 \leq \ell \leq \gamma(j, g)$$

potentielle Eltern von (j, g) , dies wird in Abbildung A.3 illustriert.

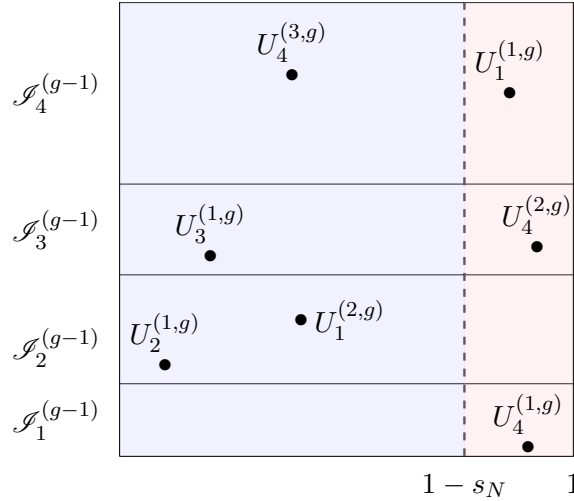


ABBILDUNG A.3: Eine Realisierung der potentiellen Vorfahren in einem Cannings-Modell mit $N = 4$ Individuen. Die Individuen 1 und 4 haben zwei beziehungsweise 3 potentielle Eltern also $\gamma(1, g) = 2$ und $\gamma(4, g) = 3$. Die Individuen 3 und 4 haben insgesamt 3 potentielle Eltern

Da wir mit a Individuen begonnen haben, erhalten wir durch dieses Verfahren a verschiedene Mengen potentieller Eltern. Wir verfolgen nur die Gesamtzahl der potentiellen Eltern, dies entspricht der Anzahl Streifen die getroffen wurden. Damit erhalten wir den oben erwähnten Verschmelzungsschritt, bei dem H Kugeln zufällig in N Kästen mit den Gewichten (W_1, \dots, W_N) verteilt werden.

Es ist zu beachten, dass eine Stichprobe der Größe n genau dann vom Wildtyp ist, falls alle potentiellen Vorfahren vom Wildtyp sind. Anders ausgedrückt, ein einzelnes Individuum ist vom bevorzugten Typ, wenn mindestens ein potentieller Vorfahre vom bevorzugten Typ ist. Diese Beobachtung ist die Erklärung für die folgende Sampling-Dualität, die die beiden Prozesse $\mathcal{K} = (K_g)_{g \in \mathbb{N}_0}$ und $\mathcal{A} = (A_g)_{g \in \mathbb{N}_0}$ verbindet.

Satz A.1.6. Sei $g \geq 0$ und $k, n \in [N]$, es gilt die folgende Dualität

$$\mathbb{E} \left[\frac{K_g(K_{g-1}) \cdots (K_g - n + 1)}{N(N-1) \cdots (N-n+1)} \middle| K_0 = k \right] = \mathbb{E} \left[\frac{k(k-1) \cdots (k - A_g + 1)}{N(N-1) \cdots (N - A_g + 1)} \middle| A_0 = n \right].$$

Der Fall $k = N - 1$ und $n = N$ in Satz A.1.6 liefert uns einen hilfreichen Ausdruck für die Überlebenswahrscheinlichkeit, welcher im Beweis von Theorem A.1.8 Verwendung findet.

Korollar A.1.7. Sei A_{eq} so verteilt wie die stationäre Verteilung des Cannings anzestral Selektionsprozesses, dann gilt

$$\pi_N := \lim_{g \rightarrow \infty} \mathbb{P}(K_g = 0 | K_0 = N - 1) = \mathbb{E} \left[\frac{A_{eq}}{N} \right].$$

A.1.4 Der Fall moderat schwacher Selektion

Wir sagen das die Selektionsstärke $s_N \in (0, 1)$ eines Cannings Modells *moderat schwach* ist, falls

$$N^{-1+\eta} \leq s_N \leq N^{-\frac{1}{2}-\eta}, \quad (\text{A.3})$$

für ein $\eta > 0$ gilt. Im Folgenden schreiben wir oft $b = b_N := -\frac{\log s_N}{\log N}$, wobei die Bedingung (A.3) dann $\frac{1}{2} < b < 1$ impliziert.

Im Verlauf dieses Abschnittes nehmen wir an, dass für ein $\rho^2 \geq 1$ gelte

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^2 \right] = \frac{\rho^2}{N^2} + O(N^{-3}), \quad (\text{A.4})$$

für $N \rightarrow \infty$. Aufgrund der Annahme (A.4) erhalten wir für die Varianz der Nachkommenverteilung

$$\text{Var} \left(\nu_1^{(N)} \right) = \rho^2 + o(1), \quad \text{für } N \rightarrow \infty.$$

Das Hauptresultat in Kapitel 3 beweisen wir unter den Bedingungen B1 sowie (alternativ) unter den Bedingungen B2.

Bedingungen B1: Wir treffen die folgenden Annahmen an das dritte Moment von $W_1^{(N)}$ und schränken das Regime der moderat schwachen Selektion weiter ein

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^3 \right] = O(N^{-3}), \quad N^{-1+\eta} \leq s_N \leq N^{-\frac{2}{3}-\eta},$$

für ein $\eta > 0$.

Alternativ dazu nehmen wir unter den **Bedingungen B2** an, dass eine Folge natürlicher Zahlen h_N existiert, so dass

$$h_N \rightarrow \infty \text{ und } h_N = o(\log N), \quad \text{für } N \rightarrow \infty,$$

mit der Eigenschaft

$$\mathbb{E} \left[\left(W_1^{(N)} \right)^n \right] \leq \left(\frac{Kh_N}{N} \right)^n,$$

für N groß genug, eine beliebige Konstante $K > 0$ und alle $n \leq 2h_N$.

Sowohl unter den Bedingungen B1, als auch unter B2, folgt, dass die reskalierte neutrale Genealogie des Cannings Modells gegen den Kingman Koaleszenten konvergiert [Möh00].

Satz A.1.8 (Haldane's-Formel, moderat schwache Selektion).

Sei $(K_g^{(N)})_{g \geq 1}$ eine Folge von Cannings Anzahl Prozessen mit Parametern N, s_N und $\mathcal{L}(\mathcal{W}^{(N)})$. Wir nehmen an, das (A.4) gilt. Unter den Bedingungen B1 oder B2, folgt für die Fixationswahrscheinlichkeit π_N eines einzelnen bevorzugten Individuums, dass

$$\pi_N \sim \frac{2s_N}{\rho^2} \quad \text{für } N \rightarrow \infty.$$

Für eine Erläuterung des Beweises verweisen wir auf den Abschnitt 2.2.3 oder auf Abschnitt 3.3.2.

A.1.5 Der Fall moderat starker Selektion

Ähnlich wie im Abschnitt A.1.4 behandeln wir den Fall moderater Selektion, jetzt im Fall in dem die Selektion stärker ist als in Abschnitt A.1.4. Wir nehmen an, dass ein $\eta > 0$ existiert, so dass

$$N^{-\frac{1}{2}+\eta} \leq s_N \leq N^{-\eta}, \quad (\text{A.5})$$

gilt, was an den Fall moderat schwacher Selektion (A.3) anschließt. Mit der vorher eingeführten Schreibweise impliziert dies $0 < b < \frac{1}{2}$.

Definition A.1.9 (Dirichlet-Typ Gewichte).

Wir nennen einen Vektor zufälliger Gewichte $\mathscr{W}^{(N)}$ vom Dirichlet-Typ, falls

$$W_i^{(N)} = \frac{Y_i}{\sum_{j=1}^N Y_j}, \quad i \in [N],$$

wobei Y_1, \dots, Y_N unabhängige Kopien einer Zufallsvariablen Y mit $\mathbb{P}(Y > 0) = 1$ sind.

Falls man Y gamma-verteilt wählt, so ist \mathscr{W} symmetrisch Dirichlet-verteilt. Das klassische Wright-Fisher-Modell erhalten wir, falls $Y = c$ fast sicher gilt, für ein $c > 0$.

Das Hauptresultat in Kapitel 4 wird unter der folgenden Annahme bewiesen

$$\mathbb{E}[\exp(hY)] < \infty, \quad (\text{A.6})$$

für ein $h > 0$. Für eine mögliche Abschwächung dieser Bedingung verweisen wir auf Bemerkung 4.3.2 a) und 4.3.3 a); im Fall einer Verteilung mit langsam variierenden schweren Rändern beispielsweise, kann man Annahme (A.6) auf die Existenz des vierten Moments abschwächen. Auch hier ist es so, dass (A.6) die Konvergenz der reskalierten neutralen Genealogie gegen Kingmans Koaleszenten impliziert, siehe Lemma 4.4.2. Unter diesen Bedingungen lässt sich der folgende Satz beweisen, der das gleiche asymptotische Verhalten der Fixationswahrscheinlichkeit wie in Satz A.1.8 beschreibt, jedoch unter leicht anderen Annahmen.

Satz A.1.10 (Haldane's-Formel, moderat starke Selektion).

Sei $(K_g^{(N)})_{g \geq 1}$ eine Folge von Cannings Anzahlprozessen mit Parametern N, s_N und $\mathcal{L}(\mathscr{W}^{(N)})$ und wir nehmen an, dass die zufälligen Gewichte vom Dirichlet-Typ seien und (A.6) gelte. Außerdem soll (A.5) erfüllt sein. Dann gilt für die Fixationswahrscheinlichkeit π_N eines einzelnen bevorzugten Individuum die Asymptotik

$$\pi_N \sim \frac{2s_N}{\rho^2} \quad \text{für } N \rightarrow \infty.$$

Der Beweis von Satz A.1.10, sowie die Bedeutung der Bedingung $0 < b < \frac{1}{2}$ werden in Abschnitt 2.2.4 genauer diskutiert.

A.1.6 Moderate Selektion und Nachkommenverteilungen mit schweren Rändern

Die Analyse von Cannings-Modellen mit moderater Selektion schließen wir mit einem Ausblick auf ein in Vorbereitung befindliches Manuskript zusammen mit M. Birkner, I. Dahmer und C. Pokalyuk ab. Dort wird die Fixationswahrscheinlichkeit eines leicht selektiven Mutanten in Cannings-Modellen betrachtet, bei denen die reskalierte neutrale Genealogie zu einer Klasse von Λ -Koaleszenten, den Beta(2 – α, α)-Koaleszenten, konvergiert, mit $\alpha \in (1, 2)$.

Definition A.1.11 (Λ -Koaleszent).

Ein Λ -Koaleszent ist ein Markov-Prozess $(X_t, t \geq 0)$ dessen Zustandsraum die Partitionen von \mathbb{N} sind und für den die Einschränkung X_t^n auf $[n]$ ebenfalls ein Markov-Prozess ist. Falls $X_t^n = k$ so findet das nächste Verschmelzungsereignis, bei dem $j \leq k$ Blöcke zu einem verschmelzen, mit Rate $\lambda_{k,j}$ unabhängig von n statt, mit

$$\lambda_{k,j} = \int_0^1 x^{j-2}(1-x)^{k-j} \Lambda(dx),$$

für ein endliches Maß Λ auf $[0, 1]$.

Wählt man für Λ das Dirac Maß in 0, so erhält man den Kingman Koaleszenten. Eine gut untersuchte Teilklasse von Λ -Koaleszenten sind die Beta-Koaleszenten, indiziert mit dem Parameter $\alpha \in [1, 2)$. Hierbei wird Λ als die Beta-Verteilung mit Parameter $(2 - \alpha, \alpha)$ gewählt, also

$$\Lambda(dx) = \beta_{2-\alpha, \alpha}(dx) := \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1} dx.$$

Hervorzuheben ist hier der Fall, $\alpha = 1$ welcher dem Bolthausen-Sznitman Koaleszenten entspricht ([BS98] und [Rue87]).

Wie im Abschnitt A.1.2 betrachten wir Cannings-Modelle die eine Paintbox-Konstruktion erlauben und wir nehmen wieder an, dass die Gewichte \mathscr{W} vom Dirichlet-Typ sind. Im Gegensatz zu der Annahme in Abschnitt A.1.5 nehmen wir jetzt jedoch an, dass die Verteilung der Zufallsvariable Y schwere Ränder mit Parameter α hat, also

$$\mathbb{P}(Y > y) \sim \ell_\alpha y^{-\alpha}, \quad \text{für } y \rightarrow \infty, \quad (\text{A.7})$$

mit $\alpha \in (1, 2)$ und einer Konstanten $\ell_\alpha > 0$.

Aufgrund des Satzes 1 iii) von Huillet und Möhle in [HM21] erhalten wir, dass die neutrale Genealogie des Cannings-Modells gegen einen $\beta_{2-\alpha, \alpha}$ -Koaleszent konvergiert (vgl. [Sch03]), wenn wir die Zeit mit dem Faktor c_N^{-1} beschleunigen, mit

$$c_N \sim \frac{\alpha B(2-\alpha, \alpha)}{\mu^\alpha} \frac{\ell_\alpha}{N^{\alpha-1}},$$

wobei $\mu = \mathbb{E}[Y]$ und $B(2-\alpha, \alpha)$ die Beta-Funktion bezeichne. Wie in den vorherigen Abschnitten betrachten wir das Regime der moderaten Selektion, also $s_N \sim N^{-b}$ und nehmen an

$$b < \alpha - 1, \quad \text{also} \quad \frac{s_N}{c_N} \rightarrow \infty, \quad \text{für } N \rightarrow \infty. \quad (\text{A.8})$$

Satz A.1.12. *Unter den Annahmen (A.7) und (A.8), gilt für die Fixationswahrscheinlichkeit π_N eines einzelnen bevorzugten Mutanten die Asymptotik,*

$$\pi_N \sim c_\alpha^{-\frac{1}{\alpha-1}} s_N^{\frac{1}{\alpha-1}}$$

mit $c_\alpha = \frac{\ell_\alpha \Gamma(2-\alpha)}{(\alpha-1)\mu^\alpha}$.

Eine kurze Bemerkung zum Beweis von Satz A.1.12 findet sich in Abschnitt 2.2.5, außerdem wird in Bemerkung 2.2.8 genauer auf die obige Asymptotik eingegangen, welche nicht mehr der klassischen Haldane Asymptotik entspricht, wie wir sie in den vorherigen Abschnitten dieser Arbeit gesehen haben.

A.1.7 Haldanes Asymptotik für Verzweigungsprozesse in zufälliger u.i.v. Umgebung

Abschließend gehen wir in diesem Abschnitt auf die Resultate und Beweistechniken, die in [BK21] verwendet werden, ein. Wir untersuchen Haldanes Asymptotik für superkritische Verzweigungsprozesse in einer unabhängig und identisch verteilten (u.i.v.) zufälligen Umgebung. Es wird gezeigt, dass Haldanes Asymptotik auch für Verzweigungsprozesse in einer zufälligen Umgebung gilt, sofern die durch die Umgebung erzeugte Fluktuationen im Vergleich zum erwarteten Drift nicht zu groß sind. Dies ist die Aussage von Satz 5. Für den Beweis haben wir uns eines bisher nicht beschriebenen Zusammenhangs mit "perpetuities" (ewigen Renten) bedient.

Zunächst definieren wir einen Verzweigungsprozess in zufälliger Umgebung und führen die notwendige Notation ein, für eine Einführung und weitere Hintergründe zu Verzweigungsprozessen in zufälliger oder variierender Umgebung verweisen wir auf [KV17] und [VZ93].

Wir bezeichnen mit $\mathcal{P}(\mathbb{N}_0)$ den Raum der Wahrscheinlichkeitsmaße auf \mathbb{N}_0 . Im Folgenden identifizieren wir jedes Maß f auf $\mathcal{P}(\mathbb{N}_0)$ mit seiner wahrscheinlichkeits-erzeugenden Funktion $f(t), t \in [0, 1]$,

$$f(t) = \sum_{k=0}^{\infty} t^k f[k], \quad t \in [0, 1],$$

wobei $f[k]$ die Masse von f in k bezeichnet. Für Wahrscheinlichkeitsmaße auf \mathbb{N}_0 kann das erste und zweite faktorielle Moment durch die erzeugende Funktion ausgedrückt werden, mit

$$f'(1) = \sum_{z=1}^{\infty} z f[z], \quad f''(1) = \sum_{z=2}^{\infty} z(z-1) f[z]. \quad (\text{A.9})$$

Statten wir $\mathcal{P}(\mathbb{N}_0)$ mit der Totalvariationsmetrik und der induzierten Borel- σ -Algebra aus, so erlaubt dies ein zufälliges Maß F auf $\mathcal{P}(\mathbb{N}_0)$ zu betrachten. Eine Folge $V = (F_1, F_2, \dots)$ von zufälligen Maßen auf $\mathcal{P}(\mathbb{N}_0)$ wird eine *zufällige Umgebung* genannt. Weiterhin nennen wir V eine u.i.v. zufällige Umgebung, wenn V aus unabhängigen und identisch verteilten $(F_i, i \geq 1)$ besteht, also die F_i unabhängige Kopien eines zufälligen Maßes F sind.

Definition A.1.13 (Verzweigungsprozess in zufälliger Umgebung).

Sei $V = (F_1, F_2, \dots)$ eine zufällige Umgebung. Wir nennen $(Z_n, n \geq 0)$ einen Verzweigungsprozess in zufälliger Umgebung, falls Z_n die Rekursion

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{i,n}, \quad Z_0 = 1$$

erfüllt, wobei $(\xi_{i,n}, i \geq 1, n \geq 1)$ bedingt auf V unabhängig sind und für alle $n \in \mathbb{N}$ die Familie $(\xi_{i,n}, i \geq 1)$ bedingt auf V unabhängig und identisch verteilt mit Verteilung F_n ist. Also sind die $\xi_{i,n}$ Kopien einer Zufallsvariablen ξ .

Wählt man $F_i = f$ fast sicher für ein $f \in \mathcal{P}(\mathbb{N}_0)$, erhält man einen klassischen Galton-Watson-Prozess mit Nachkommenverteilung f .

Vergleichbar wie in (A.9) kann das (jetzt zufällige) erste und zweite faktorielle Moment ausgedrückt werden durch

$$F'(1) = \sum_{z=1}^{\infty} zF[z], \quad F''(1) = \sum_{z=2}^{\infty} z(z-1)F[z].$$

Wir betrachten eine Folge $(Z^{(N)}, N \geq 1)$ von Verzweigungsprozessen in einer u.i.v. zufälligen Umgebung und wollen das asymptotische Verhalten der Überlebenswahrscheinlichkeit π_N untersuchen. Wir kennzeichnen die Abhängigkeit von N durch einen Index bei den Wahrscheinlichkeiten oder Erwartungen, zum Beispiel in dem wir für die Überlebenswahrscheinlichkeit

$$\pi_N = \mathbb{P}_N(Z_{\infty} > 0) := \mathbb{P}(Z_{\infty}^{(N)} > 0)$$

schreiben.

Sei

$$\varepsilon_N := \mathbb{E}_N[F'(1)] - 1, \quad \nu_N := \text{Var}_N(F'(1)),$$

also bezeichnet ε_N die Differenz von $\mathbb{E}[\xi]$ zu 1 und ν_N die Varianz des Erwartungswertes der Nachkommenverteilung zwischen den Generationen. Da uns die Überlebenswahrscheinlichkeit leicht superkritischer Verzweigungsprozesse interessiert, nehmen wir $\varepsilon_N > 0$ an und betrachten den Grenzwert $\varepsilon_N \rightarrow 0$ für $N \rightarrow \infty$. Weiterhin nehmen wir an, dass sich die Varianz von ξ stabilisiert, und setzen gleichmäßig beschränkte 4. Momente von ξ voraus. Konkret heißt das

$$\text{Var}_N(\xi) = \sigma^2 + o(1), \quad \mathbb{E}_N[\xi^4] = O(1), \quad (\text{A.10})$$

für $\sigma^2 > 0$. Außerdem nehmen wir für das erste Moment von F an, dass

$$\mathbb{E}_N[F'(1)^{-4-\delta}] = O(1), \quad \mathbb{E}_N[|F'(1) - \mathbb{E}_N[F'(1)]|^{4+\delta}] = O(\nu_N^{2+\frac{\delta}{2}}) \quad (\text{A.11})$$

für ein $\delta > 0$.

Satz A.1.14 (Asymptotische Überlebenswahrscheinlichkeit von Verzweigungsprozessen in zufälliger Umgebung).

Wir nehmen $\varepsilon_N \rightarrow 0$ für $N \rightarrow \infty$ an. Unter den Bedingungen (A.10) und (A.11) gilt:

a) Falls $\nu_N = o(\varepsilon_N)$, dann erfüllt die Überlebenswahrscheinlichkeit die Asymptotik

$$\pi_N \sim \frac{2\varepsilon_N}{\sigma^2} \quad \text{für } N \rightarrow \infty.$$

b) Falls $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$, mit $0 < \rho < 2$, dann gilt

$$\pi_N \sim \frac{(2-\rho)\varepsilon_N}{\sigma^2} \quad \text{für } N \rightarrow \infty.$$

c) Falls $\frac{\nu_N}{\varepsilon_N} \rightarrow \rho$ mit $2 < \rho < \infty$, dann gilt für große N

$$\pi_N = 0.$$

Der Beweis von Satz A.1.14 beruht auf der folgenden Proposition, die einen Ausdruck für die Überlebenswahrscheinlichkeit für jeden superkritischen Verzweigungsprozess in einer u.i.v. zufälligen Umgebung unter Bedingungen liefert, die nur geringfügig stärker sind als die üblichen Bedingungen für Superkritikalität, also $\mathbb{E} [\log F'(1)] > 0$ [SW69]. Ein ähnlicher Ansatz wurde verwendet, um die Asymptotik der Überlebenswahrscheinlichkeit im subkritischen Fall [GK00] zu untersuchen. Zu beachten ist, dass dieser Ansatz nicht die übliche Darstellung von π als Fixpunkt der wahrscheinlichkeitserzeugenden Funktion verwendet.

Für jede wahrscheinlichkeitserzeugende Funktion F mit positivem endlichem Erwartungswert $F'(1) > 0$ führen wir die *shape-Funktion* $\varphi(t)$ über die Gleichung

$$\frac{1}{1-F(t)} = \frac{1}{F'(1)(1-t)} + \varphi(t), \quad t \in [0, 1).$$

ein. Die shape-Funktion kann mithilfe einer Taylorentwicklung stetig auf das Intervall $[0, 1]$ fortgesetzt werden, durch

$$\varphi(1) = \frac{F''(1)}{2F'(1)^2}.$$

Im Folgenden bezeichnen wir mit $\varphi_k(t)$, $t \in [0, 1]$ die zu F_k gehörige shape-Funktion in Generation $k \in \mathbb{N}$. Außerdem schreiben wir

$$\mu_k = \prod_{i=1}^k F'_i(1), \quad \mu_0 = 1.$$

Proposition A.1.15. *Sei $(Z_n, n \geq 0)$ ein Verzweigungsprozess in u.i.v. zufälliger Umgebung $V = (F_1, F_2, \dots)$, welche aus unabhängigen Kopien von F besteht. Sei*

$$0 < \mathbb{E} [\log F'(1)] < \infty \text{ und } \mathbb{E} [\log^+ F''(1)] < \infty.$$

Dann kann die Überlebenswahrscheinlichkeit π ausgedrückt werden durch

$$\pi = \mathbb{E} \left[\frac{1}{X} \right],$$

mit

$$X := \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(\mathbb{P}(Z_{\infty} = 0 \mid Z_{k+1} = 1, V))}{\mu_k} < \infty \quad f.s.$$

Mit Proposition A.1.15 bleibt zum Beweis von Satz A.1.14 noch die Asymptotik von $\mathbb{E} \left[\frac{1}{X} \right]$ unter den Annahmen von Theorem A.1.14 zu bestimmen. Da wir uns im leicht superkritischen Fall befinden, scheint es vernünftig zu erwarten, dass die Aussterbewahrscheinlichkeit für $N \rightarrow \infty$ gegen 1 konvergiert, also $\mathbb{P}_N(Z_{\infty} = 0 \mid Z_{k+1} = 1, V) \rightarrow 1$ für $N \rightarrow \infty$. Daher betrachten wir die approximierende Zufallsvariable Y von X , definiert durch

$$Y := \sum_{k=0}^{\infty} \frac{\varphi_{k+1}(1)}{\mu_k}.$$

Man beachte, dass Y die stochastische Rekursion

$$Y = \varphi_1(1) + \sum_{k=1}^{\infty} \frac{\varphi_{k+1}(1)}{\mu_k} \stackrel{d}{=} \varphi(1) + \frac{1}{F'(1)}Y, \quad (\text{A.12})$$

erfüllt, wobei $\varphi(1)$ und $F'(1)$ unabhängig von Y sind. Die Rekursion (A.12) wird im finanzmathematischen Kontext manchmal *Annuitäts-Gleichung* genannt [BDM16] und entsprechend Y als *Perpetuity* (ewige Rente) bezeichnet. Für Zufallsvariablen Y die (A.12) erfüllen mit nicht negativen Zufallsvariablen (A, B) , so dass

$$Y \stackrel{d}{=} A + BY,$$

mit

$$\beta_N := 1 - \mathbb{E}_N[B], \quad \gamma_N := \text{Var}_N(B),$$

beweisen wir den folgenden Grenzwertsatz für ewige Renten mit verschwindendem Zinssatz, welcher auch unabhängig von dem Kontext dieser Arbeit von Interesse sein kann.

Satz A.1.16 (Grenzwertsatz für ewige Renten).

Angenommen, es gelte im Grenzwert $N \rightarrow \infty$, dass

$$\beta_N \rightarrow 0, \quad \gamma_N \rightarrow 0, \quad \mathbb{E}_N[A] = \alpha + o(1), \quad \mathbb{E}_N[A^{1+\delta}] = O(1),$$

mit $\alpha > 0$ und ein $\delta > 0$. Außerdem gelte $\mathbb{E}_N[|B - 1|^{2+\delta}] = o(\beta_N + \gamma_N)$, sowie

$$\frac{\beta_N}{\gamma_N} \rightarrow \hat{\rho} \quad \text{mit} \quad -\frac{1}{2} < \hat{\rho} \leq \infty.$$

Dann gilt:

- i) Im Fall $\hat{\rho} = \infty$, konvergiert $\beta_N Y$ gegen α in Wahrscheinlichkeit.
- ii) Im Fall $\hat{\rho} \in (-1/2, \infty)$, ist $\gamma_N Y$ asymptotisch *invers-gamma* verteilt, mit Dichte $\frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x} dx$ auf \mathbb{R}^+ und Parametern $(a, b) = (2\hat{\rho} + 1, 2\alpha)$.

Der Beweis von Satz A.1.14 baut auf dem obigen Satz auf, für weitere Details verweisen wir an dieser Stelle auf Abschnitt 2.3.