# Cultural Transmission of Fitness in Mathematical Population Genetics



D.Phil. Thesis

Imre Kvasznicza

Hertford College

in the University of Oxford

Trinity Term, 2014

# Cultural Transmission of Fitness in Mathematical Population Genetics

Imre Kvasznicza, Hertford College, D.Phil., Trinity Term 2014

Cultural transmission of fitness is the observed phenomenon in human populations that the number of a parent's siblings influences the number of their own offspring. We propose an extension of the classical Wright-Fisher model for constant size haploid populations to allow for setting the magnitude of these 'influences' for each family size separately. This is achieved by assigning a reproductive weight to each individual in the parental generation depending on their family size.

The model is a robust extension in the sense that both the forward-in-time model and its dual coalescent process behave similarly to Wright-Fisher and the latter's own limit dual, the $n$-coalescent, respectively, provided the weight sequence has a constant tail, that is, individuals with many siblings receive the same reproductive weight. Indeed, in the limit of infinite population size, the coalescent dual of our model converges weakly to a constant scaled $n$-coalescent.

Second, we present results of numerical simulations of the extended Wright-Fisher model with strictly increasing weight sequences of the form $w = (j^\alpha)$, $\alpha \geq 0$. Here we observe a phase transition in family size frequency at the point $\alpha = 1$; and we report multiple mergers occurring in the model's coalescent dual. We also use the data to approximate the coalescent's natural time scale. Such a rapidly increasing weight sequence as this one seems to lack biological relevance, however.

Third, we study a near-critical multi-type branching process model of the phenomenon. A sequence of rescaled Galton-Watson branching processes weakly converges to a Feller-type diffusion. The drift and the diffusion coefficients are expressed in terms of the reproductive weight sequence.

Fourth, we describe a branching diffusion model for cultural transmission of fitness whose dual process is the Bolthausen-Sznitman $U$-coalescent demonstrating that the phenomenon can naturally lead to a whole variety of dual processes with different qualitative features.

# Contents

**1 Introduction** **1**

1.1 Chequered history . . . . . 7

1.2 Correlation in pedigrees . . . . . 11

1.3 What shapes the family size? . . . . . 14

**2 Three models: two for simulation, one for analysis** **17**

2.1 Austerlitz and Heyer's branching process model . . . . . 17

2.2 The Wright-Fisher model . . . . . 18

2.3 Sibert, Austerlitz and Heyer's extension of the Wright-Fisher model . 21

2.4 A tractable model for cultural transmission of fitness . . . . . 22

**3 Part I: An extension of the Wright-Fisher model to cultural transmission of fitness** **26**

3.1 Convergence in population size . . . . . 26

3.2 Convergence in time . . . . . 35

3.3 The impact of cultural transmission on family size . . . . . 51

3.4 The coalescent dual of the model . . . . . 62

3.5 Conclusion . . . . . 85

**4 Part II: Computer simulations** **88**

4.1 Family sizes . . . . . 91

4.2 Number of reproducing daughters . . . . . 99

4.3 Tree topology . . . . . 100

4.4 Conclusion . . . . . 108

**5 Part III: A multi-type branching process for cultural transmission of fitness 110**

5.1 The mean and the variance matrices, the Perron root and the long time behaviour of the process . . . . . 114

5.2 The mean operator: the infinite-dimensional case . . . . . 124

5.3 Diffusion approximation of family and allele frequencies . . . . . 138

5.4 An extension with weight sequences of the form $w = (1 + j\alpha/N)$ . . . . . 147

5.5 Conclusion . . . . . 163

**6 Part IV: A branching diffusion with unlimited different family size types 165**

6.1 Derivation of the underling spatial motion . . . . . 166

6.2 The coalescent dual of the model . . . . . 169

6.3 Concluding remarks . . . . . 172

**7 Summary and future directions of research 174**

**A Sufficient conditions for existence of a time-stationary family size distribution in Part I 184**

**B Parameters for the approximations in Part II 187**

**C Matlab code for Part II 188**

# 1 Introduction

The change in allele frequency in a population is traditionally explained by four forces acting on the alleles: genetic drift, selection, mutation and migration. The first two phenomena decrease genetic diversity, but they do so in different ways. Genetic drift is the result of random sampling from the previous generation; it decreases variation because some alleles do not spawn offspring. The effect of selection, however, is not uniform on the whole population: an advantageous allele in a panmictic population spreads on average at a logistic rate determined by the size of advantage. Mutation and migration offset these two mechanisms by introducing new variants into the population, and thus enhancing its diversity.

Over the past seventy years this small set of forces was used to describe the change in allelic frequencies in many laboratory populations surprisingly accurately. For example, using the most popular model, Buri (1956) successfully predicted the variance in eye colour of a *Drosophila melanogaster* population.

The past decades were witness to population genetics' enormous success in reconstructing genealogies in a given species. The models of generation-by-generation evolution are rescaled to give rise to various coalescent processes. By looking at observed genetic data at their foot, these coalescent processes are then used to infer the history of the species. The underlying models hence need to be reasonably simple so that one can pass to a coalescent limit. This tractability is often achieved by assuming that offspring numbers are exchangeable. In human populations, several accounts have refuted the assumption of exchangeability, however.

Austerlitz and Heyer (1998) observed that the peculiar pattern of inherited diseases such as cystic fibrosis and muscular dystrophy in the Saguenay-Lac-St-Jean region, Québec, Canada, cannot be explained by the usual mathematical description

of genetic drift. (The complete list of the identified pathogenic alleles appears in Scriver (2001).) The genetic drift in the well-studied Wright-Fisher model is blind: each allele has equal chance of having offspring. Thus the effective family sizes in the subsequent generations in Québec should be uncorrelated. When the authors took the genealogical records they saw that, in fact, there was a trans-generational correlation between family sizes. In order to capture this phenomenon, Austerlitz and Heyer (1998) generalised the mechanism by which the new generation is obtained from the previous one so that the number of an individual's siblings influences the number of the individual's offspring. In their model it is possible to observe correlation. Austerlitz and Heyer call this new version of drift a model for the cultural transmission of fitness (abbreviated CTF). We describe in detail both this and the basic Wright-Fisher model in §2.

After having given a brief overview of the phenomenon, we propose three simple models for CTF in §2. Several models aiming to emulate CTF have been introduced in the literature, yet these are difficult to analyse and rather excessive in their biological assumptions. Our constructions are both mathematically accessible and flexible enough to fit family size distributions characteristic of human populations. In Parts I (§3) and II (§4), we study an extension of the classic Wright-Fisher model for CTF. Here the size of the haploid population is assumed to remain constant throughout the generations, thus the model captures the competition among the individuals. However, in many rapidly expanding populations, where CTF is believed to be present, such as among the inhabitants of the Saguenay-Lac-St-Jean region or the Hutterites in America, individuals do not compete for resources. The evolution of these populations is better captured by a branching process. This case is discussed in Parts III (§5) and IV (§6). Below we lay out a more refined structure of the thesis, and state our main results.

Cultural transmission of fitness is the phenomenon that the number of an individual's siblings influences the number of her own offspring. In §2.4 we propose an extension of the classical Wright-Fisher model for haploid populations of constant population size $N$ to allow for setting the magnitude of these "influences" for each family size separately. This is achieved by assigning a reproductive weight to each individual in the parental generation depending on their family size: an individual of family size $i$ receives the weight $w_i \in \mathbf{R}^+$. This individual will be chosen as the mother of any given individual in the new generation with probability $w_i / \sum_{j=1}^{N} N p_j w_j$, where $p_j$ stands for the proportion of individuals of family size $j$ in the parental generation. Thus our extension belongs to the general class of frequency-dependent selection models.

In §3.1, assuming that the weight sequence $w$ has a constant tail, (that is, for some $K \in \mathbf{N}^+$, we have $w_i = w_K$ for all $i \geq K$) we find a recursion that the family size distributions $(p_j)$ in subsequent generations follow, and show that, in the infinite population size limit, the family size distribution is a convex combination of Poisson distributions. In §3.3 we study the moments of, and the correlation between the family size of a randomly picked individual from the population and of its parent's family size. We also express the parameters of the model in terms of these two distributions. In §3.4 we investigate the genealogies of a sample taken from the population. For our main result in Part I, we use a collapse of structure argument to show that, in fact, the number of surviving lineages (time measured in units of the population size) converges weakly to a rescaled $n$-coalescent: an ancestral process in which only pairs of lineages coalesce.

**Theorem 1.1.** *Measuring time in units of the total population size $N$, the ancestral process $A_N(\lfloor Nt \rfloor)$ of the CTF extension of the Wright-Fisher model converges weakly to the rescaled $n$-coalescent $A(ct)$, as $N$ becomes large, provided that the weight*

*sequence $w$ has a constant tail. The scaling factor $c$ is given by*

$$c = \sum_{j=1}^{\infty} \frac{j-1}{p_j} \gamma_j^2,$$

*in terms of the time-stationary family size distribution $(p_j)$ and the invariant measure $(\gamma_j)$ of the backward migration process of the lineages.*

A single lineage 'migrates' from a family size class to an other family size class in each generation as time passes backwards. We call this the backward migration process of the lineages above. In §3.4 we give a method to calculate its invariant measure $\gamma$.

The theorem requires that the family size distributions $(p_j(t))$ converge in time. The dependence of this convergence on the choice of the weight sequence $w$ is investigated in §3.2. Our conclusion here is that family size reaches a time stationary distribution whenever the weights $(w_i)$ are 'close enough' to each other. We conjecture that the constant tail assumption of the sequence $w$ can be weakened: uniform boundedness of $w$ in $N$ should be sufficient for the convergence theorem to hold.

In Part II, we present results of numerical simulations of the above extended Wright-Fisher model with strictly increasing weight sequences of the form $w = (j^\alpha)$, $\alpha > 0$. Here we observe a phase transition in family size frequency at the point $\alpha = 1$; and we report multiple mergers occurring in the model's coalescent dual. We also use the data to approximate the coalescent's natural time scale. Such a rapidly increasing weight sequence as this one seems to be very strong, lacking biological relevance. We chose to investigate this case to mimic Sibert et al. (2002).

Part III discusses a multi-type Galton-Watson process model for cultural transmission of fitness. A branching process was proposed by Austerlitz and Heyer (1998) to capture CTF, but here we work with the branching process version of our ex-

tended Wright-Fisher model. Individuals are again classified by their family size but they reproduce independently; the offspring number of an individual of family size $i$ has probability mass function $(p_j(w_i))$ with mean $w_i$ and finite variance. Choosing the $w$'s to be close to one, we make the branching process near-critical. In §5.2 we calculate the principal eigenvalue of the mean matrix of this branching process. Using the principal eigenvalue and the corresponding eigenvector, we write down the generator of the Feller diffusion limit of the rescaled process. As a consequence, we are able to describe the evolution of allele frequencies in a population evolving according to this branching process model. Each individual in the population carries one of the two different alleles ($A$ or $a$). This allele is then transmitted to her offspring. Assuming that mutation is not present, we obtain the following theorem.

**Theorem 1.2.** *Measuring time in units of $N$, the proportion of $A$-type alleles converges weakly, as $N \to \infty$, to the diffusion $q_t$ given by*

$$dq_t = \sigma \sqrt{\frac{q_t(1-q_t)}{s_t}}\, dB_t$$
$$ds_t = \beta s_t\, dt + \sigma \sqrt{s_t}\, dB_t,$$

*and stopped if $q_t$ hits 0 or 1, in other words, if the allele $A$ fixes or dies out, respectively. Here $s_t$ corresponds to the total size of the population, and $B_t$ is a standard Brownian motion. The constants $\beta$ and $\sigma$ are determined by*

$$\beta = \sum_{i=1}^{\infty} i\epsilon_i p_i(1) \text{ and}$$
$$\sigma^2 = \sum_{i=1}^{\infty} i(i-1)p_i(1),$$

*where an individual of family size $i$ has offspring distribution $p(w_i)$ with $w_i = 1 +$*

$\epsilon_i/N$.

Of course, this result corresponds to the familiar case where the population evolves according to a single-type near-critical Galton-Watson process with offspring number mean $1 + \beta/N$ and variance $\sigma^2$. Our contribution here is the calculation of the explicit value of $\beta$ in the multi-type context of cultural transmission of fitness.

Part IV describes a subcritical branching diffusion whose underlying motion is chosen in such a way that we can interpret the particles' spatial position as their family size. In this model for CTF, greater family size is associated with strictly greater reproductive fitness, reflecting the assumption that a mother's benefit from cooperation with her siblings strictly increases with their numbers without bound. The spatial motion has a negative drift term modelling the assumption that the fertility of an ageing mother slowly decreases in average over time, or indeed that the number of elder siblings has a detrimental effect on the number of an individual's yet to be born children. We shall see that here the genealogies qualitatively differ from the ones in the first two models. In §6.2 we apply Theorem 3 in Berestycki et al. (2013) to deduce that the model's limit dual is a time-scaled $U$-coalescent: multiple mergers of the lineages occur. This demonstrates the notion of CTF can lead out from the familiar framework not only in the case of a constant size population (see Part II) but also of branching process models, as Austerlitz and Heyer (1998) predicted.

Finally, we note that *family size* is an ambiguous term, we need to be careful to specify what we mean by it throughout this thesis. We say that an individual has family size $j$ if she was born into a family of size $j$, that is, if she has $j - 1$ sisters. In particular, it does not mean that the individual has $j$ daughters.

## 1.1 Chequered history

The idea of handing down the "love for children" in families fascinated the pioneers of genetics: Pearson and Lee (1899) surveyed the trans-generational correlation in family sizes of British peers. The study of Cultural Transmission of Fitness is thus as old as Pearson's correlation coefficient. The concept was also debated in eugenic circles. A notable example is Alexander G. Bell's short article on how to avoid race suicide (Bell, 1920). He takes an island with declining population size, and speculates that those who have a single offspring, because they learned to do so from their parents, will eventually die out. When the tides turn – the inventor remarks – it is important to put strict immigration laws in place to avoid the dilution of the island race by outsiders who would typically possess a higher fertility rate relative to the average inhabitant.

The scientific examination of the relationship in family sizes of mothers and daughters dates back to the birth of modern population genetics. Ronald Fisher noted the trans-generational correlation in family sizes in the very same book, The Genetical Theory of Natural Selection (1930), §§4 and 5, from which the celebrated Wright-Fisher model emerged via Sewall Wright's work (Wright, 1931). Berent (1953) compared the family sizes of hospitalised women with the sizes of their parents' families. The correlation was present in all social strata, among mothers and fathers, as well as among birth-controllers and non-controllers. Huestis and Maxwell (1932) conducted a similar enquiry among their 638 female students at Oregon University. At the time a decrease in the population of the U.S.A., attributed to the introduction of contraception, was feared. The study suggested that the positive correlation (0.12) found in the family sizes between the two successive generations was due to the cultural handing down of the practice of contraception. "[...] as this

knowledge seeps downward from its centre of origin among the illuminati, to more and more completely film the moiling masses of the unlettered, they will become unworthy of their title proletarians, and depopulation will begin."

The simple models which became popular in the 1920s-30s do not capture cultural transmission of fitness. However, they predict human genetic history in agreement with archaeological findings. As a result, the attention turned away from CTF. It was the evidence found in the Saguenay-Lac-St-Jean study (Austerlitz and Heyer, 1998) which turned the spotlights to CTF again. Since then a myriad of new cases displaying trans-generational correlation in family sizes has come to light.

Belle et al. (2006) draws up 10 different possible scenarios for modern day Tuscans descending from the Etruscans who inhabited Italy before the Roman times. The set-ups include a small population of constant size, an expanding population and two independent populations with no migration. These are then tested against collected mitochondrial DNA data. The second most likely course of events was a single founder effect taking place about 240 generations ago. (A founder effect is an event that a small, non-random sample of individuals become separated from a larger population, decreasing the genetic variety of the new "colony", as a result. The separation is, of course, not necessarily geographical.) The most probable set-up was the same founder effect with high mutation rate. The researchers suggest that the founder effect was achieved as CTF lowered the number of reproducing alleles: the elite genetically differentiated from the rest of the society. The effect of CTF on the number of reproducing individuals is discussed in §3.3. Particularly, we shall see that a change in society might have similar consequences to a "founder effect".

Blum et al. (2006) uncovered fertility inheritance in hunter-gatherer populations. Hunter-gatherers had larger correlation in family size than cultivators. Blum argued

that individuals belonging to larger kin networks benefited from stronger social support. However, it is equally likely that early agriculture simply could not support large families. The article also developed a novel method for the detection of CTF. Before, fertility correlation was worked out by looking at the family sizes of related individuals recorded in the genealogical datasets. Instead of this practice, the authors examined the imbalance of genetic trees after reconstructing them from a sample of sequences. A vertex of the tree is said to be completely balanced if the vertex splits the subtree into two subtrees of equal size. The imbalance of a vertex increases with the difference in the sizes of the subtrees. Finally, the imbalance of the tree is the average of the imbalance of its nodes. Blum finds that tree balance characterises CTF better than trans-generational correlation when only incomplete information is available. In §3.4, we show that, in our constant population size model with weight sequences of constant tail, all genetic trees are equally likely, and hence tree imbalance statistics do not discriminate CTF. For strictly monotone increasing weight sequences of the form $w = (j^\alpha)$, the genealogical trees stay balanced for values of $\alpha \in [0, 1]$, but in the case of more rapidly increasing sequences ($\alpha > 1$), we shall see in §4.3 that the genealogies become imbalanced. In §4.4, assuming our extension of W-F model with weight sequences $w = (j^\alpha)$, we describe how to infer the value of the parameter $\alpha$ using the trans-generational correlation in family size and Blum's balance statistic.

Lastly, Chaix et al. (2008) investigates human demographic expansions in Eurasia using Y chromosome and mitochondrial DNA data. The time of the expansion is shown to decrease as we move from east to west which, in turn, suggests that demographic expansion diffused westward through migrations of individuals and not just through the spread of technologies. There was about 20 thousand year difference between expansion ages inferred from Y chromosome and expansion ages obtained

from mitochondrial DNA. The odd mutation rates might account for the imprecision but – the authors point out – the error could be explained by the different social behaviour of males and females. Accepting that fertility is transmitted from mother to daughter culturally, some difference in social behaviour among males and females may lead to mothers' lines evolving at a different pace than fathers' lines. Trans-generational transmission of fitness could have resulted in a reduction of variation in the male-transmitted Y chromosome compared to the matrilineal mitochondrial DNA. Admittedly, in Chaix et al. (2008), the high mutability of the Y chromosome is a more plausible explanation than the author's hypothesis that CTF is at work in the migrating population. In contrast, the greater impact of CTF on matrilines is confirmed by both Berent (1953); Helgason et al. (2003), see §1.2. In §3.4 we show that patrilines and matrilines do evolve at different speeds, assuming that males' social behaviour differs from females'.

All three examples above came after the publication of Heyer et al. (2005), a review on CTF. Their article enumerates the cases of CTF observed up to that point. Trans-generational correlation of family sizes had been seen in the Icelandic population; in some Amerindian tribes; and among Genghis Khan's progeny. Other studies reported on the transmission of social status in Maoris; on the Hopi Indians' transmission of albinism; and on the segregation of one part of the inhabitants in the Valserine Valley, France from the rest. CTF was perceived to be present in animal and plant populations, too: matrilineal transmission in whales; in Serengeti cheetahs; as well as in certain chimpanzee populations; and finally in some artificial species of wheat. Thus, although the name *cultural transmission of fitness* suggests that the phenomenon is present in social species only, selection based on family size may affect many more populations. Perhaps another example beside genetically modified crop is a cancer cell population, where the number of a single cell's

descendants per unit time is a good indicator of its malignity.

## 1.2 Correlation in pedigrees

Looking back only one generation in time as Berent, Huestis and Maxwell did, we cannot learn as much about transmission of fertility as if we had studied whole pedigrees. Examining the church records we may be able to answer a multitude of questions. Does the effect of CTF wear off as we traverse the pedigree backward in time? If it does, how fast do we lose information of the family size configuration from which we started? Is it possible that correlation alternates in sign through the generations? To what extent is the time to the most recent common ancestor attenuated or shortened? In turn, incorporating the observations into a new model, we should be able to address questions on the long-term consequences of the phenomenon which remain hidden in shorter pedigrees. To what coalescent do the genealogies converge when population size becomes large? What Markov process may approximate the allele frequencies in the population?

The three pedigrees available are church records of isolated communities of Canadians, Icelanders and the Hutterites stretching three or four centuries back in time. We take time to introduce these populations to our reader.

The Hutterites form a distinct branch of Anabaptism. The followers of Jakob Hutter were ousted from Catholic Austria because of their religion. After four centuries of flight, today they live in America. In their new home the 400 strong group grew to 4 200 in just under 125 years.

The Saguenay-Lac-Saint-Jean region has a population of 300 000. They descend from only 5 000 French settlers who colonised Quebec at the end of the 17th century. The region itself was uninhabited until the 19th century. The origin of the settlers is also known: the males arrived from the shore, whereas most women originated

from Lyon and the Île-de-France. Thus the founder genes are geographically rather dispersed. Today, the four most frequent disorders have carrier frequencies between 0.035 and 0.05. These must have increased from 1/5 000 because the genes were introduced as a single copy at settlement (Austerlitz and Heyer, 1998). In particular, the distribution of the number of married children per married woman can be extracted from the records (Tremblay and Vézina, 2010). The mean of the distribution is 3.4 married children per married women throughout the 20th century. In this time the correlation in the effective family size of an individual and the effective family size of her mother is determined to be 0.21. (By the effective family size of an individual we mean the number of reproducing siblings including herself; this notion is often used in the literature of CTF.) A century ago the correlation was even more pronounced: its value reached 0.34. There is little competition in finding a mate in these rapidly expanding populations so much so, that, in the Hutterite communities, marriages are matched by colony leaders (Hostetler, 1997).

Helgason et al. (2003) traces back 131 060 individuals' family histories over a 300 year period in Iceland, considering fathers' (Y chromosome) and mothers' (mitochondrial DNA) lines separately. The author finds that the matrilineal ancestors left more descendants than the patrilineal ones, and concludes that women's lines evolve at a faster rate than men's. As the same level of reproductive variance within the generations is gauged for both kinds of lines, reproductive variance does not explain the faster evolutionary rate of the matrilines. They attribute the phenomenon to cultural transmission of fitness, like Chaix et al. (2008). Indeed, the trans-generational correlation is always found to be stronger in mothers' lines than in fathers'. This agrees with the result of the two generation deep survey in Berent (1953).

Table 1 summarises the values of trans-generational correlation in family sizes

found in the human populations described above. For comparison, the last row of the table shows data regarding the Japanese population where the association between the family size of a randomly picked individual and its parent's family size is weak, while among the Hutterites the association is the most pronounced.

Table 1: Correlation between family size of an individual and of its parent

| Population | Correlation in family size | |
|---|---|---|
| | fathers-sons | mothers-daughters |
| Oregon University students | | 0.124 |
| British aristocracy | 0.066 | 0.21 |
| SLSJ region in Québec | 0.14 | 0.17 |
| Icelanders | 0.041 | 0.079 |
| Hutterites | 0.31 | 0.23 |
| Japanese | -0.031 | 0.014 |

Lastly but most importantly, a recent study (Murphy, 2007) finds that in the 1987 U.S. National Survey of Families and Households the correlation between the respondents' family size and their number of children ranges between 0.104 and 0.229 in the different age groups. (In this survey family size means the number of siblings with whom the parents were growing up.) It is fascinating that the correlation had been steadily increasing in a 40-year period as shown in Table 2.

Table 2: Correlation between number of siblings and children, by age of parent; National Survey of Families and Households 1987-8; taken from Murphy (2007)

| Age-group | U.S. National Survey 1987-8 | |
|---|---|---|
| | correlation coefficient | sample size |
| 35-44 | 0.229 | 2 731 |
| 45-54 | 0.177 | 1 528 |
| 55-64 | 0.154 | 1 356 |
| 65-74 | 0.143 | 1 200 |
| 75+ | 0.104 | 797 |

We draw three conclusions. First, the magnitude of trans-generational correlation in family size is small but detectable. Second, the correlation coefficient can

be negative. Third, the level of correlation may change between subsequent generations. We shall see in Parts I (§3.3) and II (§4.2) that our proposed model exhibits correlation with these properties.

## 1.3 What shapes the family size?

Human offspring are not drawn from a beanbag: the binomial distribution of the offspring number in the Wright-Fisher model is rarely a good approximation in the case of human populations. From a survivalist point of view, there is an optimal family size for each social species. If a family is large enough, its members benefit from cooperation. But having too many offspring can quickly become a burden, when their parents cannot feed them. This explains the peculiar shape of the family size distribution that we see in primate populations: it tends to flatten out about the optimal size as seen in Figure 1.

Huestis and Maxwell (1932) go further and suggest some factors which might have a sway on human family size. The vanishing margin of "parental desire" is best illustrated by example. Having many children already, the birth of two extra offspring does not add as much benefit as the first two did. Another element is social pressure. A woman may not have another baby, even if she would like one, just to conform with the majority who have two children. But peer pressure works in both directions: in rural Eastern Europe most Romany families bring up seven or eight children. Finally, if a child grows up in a large family, she might become mindful of the economic and social disadvantage of having too many children, and decide to have a smaller family herself. (This might cause negative correlation.)

Lastly, we discuss the effect of social status on the family size. Social standing influences child survival via access to resources. Status is very much heritable (think of university education and inheritance of assets); and this perhaps was even more

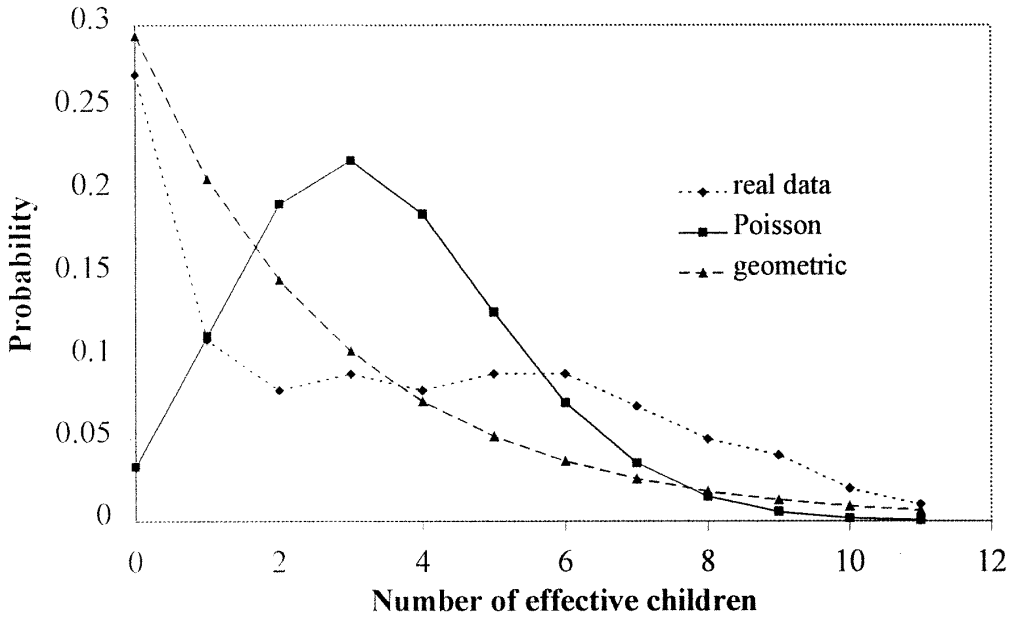



Figure 1: Distribution of effective family sizes per married woman in the 19th century Saguenay-Lac-St-Jean region in Québec compared with the Poisson and geometric distributions of the same mean. Taken from (Austerlitz and Heyer, 1998), see the definition of effective offspring number in §2.1.

so in the past. Belle et al. (2006) suggests the genetic differentiation of the Etruscan elite. It might be that the privileged members of the ancient Italian society married among each other. Yet that we see their genetic footprint today, after three millennia, implies that they must also have had many children. The claim that every third Mongolian is related to Genghis Khan is explained similarly: the overlord could not have that many children himself. It seems more plausible that his progeny had numerous offspring, as well as their great progenitor, taking advantage of their ancestry. So both low and high status individuals may hand down their number of children together with their social circumstances.

In this short review, we have seen that beside the observed correlation in family sizes, the shape of the family size distribution is also characteristic of CTF. So, whatever the causes are, a good model for demographics should allow for an inter-

generational correlation in family size while it leaves the distribution flexible. The next section shows that, in order to have a model like this, we need not cast away our beanbag after all; just change the rules of drawing.

# 2 Three models: two for simulation, one for analysis

In this chapter we look at the models for cultural transmission of fitness suggested in Austerlitz and Heyer (1998) and Sibert et al. (2002); and we briefly discuss in what circumstances they are most applicable. We also introduce our models which will be analysed in the remaining parts of the thesis. Both Sibert's and our constant population size model are extensions of the standard Wright-Fisher, so a description of this classic model is also included here for completeness. Austerlitz and Heyer's branching process is a diploid model that exhibits trans-generational correlation in family size. We instead suggest a more approachable multi-dimensional Galton-Watson process in Part III, and a branching diffusion to capture the phenomenon in Part IV.

## 2.1 Austerlitz and Heyer's branching process model

Austerlitz and Heyer (1998) simulated pedigrees in the Saguenay-Lac-Saint-Jean region in Québec using a branching process conditional on the family sizes of the parents. The process models the demographics of a growing, diploid population, and uses the following notion of effective family size.

**Definition 2.1** (Number of effective children)**.** *We define the effective family size of an individual to be the number of their reproducing children.*

**Definition 2.2** (Austerlitz and Heyer's branching process model)**.** *In generation $t$, couples are formed at random from the total population present in generation $t-1$. The effective family size of each pair is drawn from a Geometric or Poisson distribution with mean $\alpha_t[(1-c)\lambda + c(S_1+S_2)/2]$, where $S_1$ and $S_2$ are the parents'*

*effective family sizes. The parameter $\lambda$ is the rate at which the population grows, $\alpha_t$ is a normalisation parameter which guarantees that the total growth rate remains $\lambda$, while $c$ is a mixing parameter between 0 and 1.*

Thus the correlation in the family size of a randomly chosen child and the family sizes of its parents depends only on the parameter $c$. The genealogies then are simulated using the pedigrees obtained this way. Two unique alleles are assigned to each founder. Then one allele is selected from each parent to be transmitted to the child.

A branching process does not impose a constraint on the total population size and hence cancels competition amongst the individuals in finding a mate. This model therefore seems to be most suitable for populations characterised by rapid expansion and little competition like the ones introduced in §1.2 (with the exception of Icelanders, perhaps). Correlations however are present in almost all human populations, and competition is not negligible in general. Thus, we should like to capture competitive mating.

Apart from the lack of provision to emulate competition, this branching process falls short even in the targeted Saguenay-Lac-St-Jean population: as Figure 1 shows, the empirical family size distributions are neither geometric nor Poisson. Whereas, as outlined in the previous section, a control over the distribution of family sizes seems to be key in modelling cultural transmission of fitness.

## 2.2 The Wright-Fisher model

In order to compare Sibert, Austerlitz and Heyer's model (Sibert et al., 2002) with our own, we need to understand first the Wright-Fisher model (abbreviated W-F) on which the two extensions are based. In W-F, we keep track of a haploid population

of alleles through consecutive, non-overlapping generations. The population size is assumed to be constant: there is no growth or decay in the process. We also suppose that the population is neutral, that is, the alleles have equal chance to reproduce. Thus, the new generation is obtained from the previous one by choosing a parent for each allele at random with replacement. That is, consecutive generations are produced by multinomial sampling. This rule can be put formally as follows.

**Definition 2.3** (The Wright-Fisher model)**.** *Writing $N$ for the population size and $(\nu_1, \nu_2, \ldots \nu_N)$ for the number of offspring which form generation $t$ ($\nu_i$ is the number of children of the $i$th individual in the parental generation), we have $(\nu_1, \nu_2, \ldots \nu_N) \sim Mult(N;\ 1/N, 1/N, \ldots 1/N)$, in other words, the offspring number vector $\nu$ has a Multinomial distribution with parameters $N$ and $p = (1/N, \ldots\ 1/N)$.*

For our later discussion, here we note that it follows that the number of offspring of an individual has a Binomial distribution with parameters $(N, 1/N)$. This in turn can be approximated by a $Pois(1)$ distribution, when the population size is large. (We adopt the notations $Bin(n, p)$ for the Binomial distribution with number of independent trials $n$ and success rate $p$, and $Pois(\lambda)$ for the Poisson distribution with mean parameter $\lambda$.) Another easy observation is that the time to the most recent common ancestor (abbreviated MRCA) of a pair is geometrically distributed with success rate $1/N$. To see this, notice that the probability that two alleles have the same parent is $1/N$. If they do not, then the probability that their parents descend from a common parent again is $1/N$, and so on. Multiple lines coalesce with probability $O(1/N^2)$. Thus, should we measure time in units of generations the probability that two lineages remain distinct, $\left(1 - \frac{1}{N}\right)^{\lfloor Nt \rfloor}$, converges to $e^{-t}$, as $N$ becomes large. We do not see multiple mergers in the infinite population size limit, as the probability of these events, after rescaling time, is still $O(1/N)$. The

number of ancestors of a sample, in this case, is therefore given by the the following Markov chain, the limit coalescent dual of W-F. (Note that there are $\binom{i}{2}$ pairs of $i$ ancestral lines to merge with exponential rate 1.)

**Definition 2.4** (The ancestral process in the $n$-coalescent)**.** *The ancestral process or the block counting process of the $n$-coalescent, $A(t)$ is the continuous-time Markov chain tracing back in time the number of ancestors of a sample of $n$ individuals with state space $\{1, 2, 3, \dots n\}$, initial state $A(0) = n$ and transition rates $Q = (q_{ij})$, given by $q_{ii-1} = \binom{i}{2}$, $q_{ii} = -\binom{i}{2}$ and $q_{ij} = 0$ for $j \neq i, i-1$.*

However, the $n$-coalescent is meant to be more than the ancestral process. If we label individuals in our sample $\{1, 2, 3, \dots n\}$ then the system of coalescing lineages defines a continuous Markov process $\pi(t)$ on the equivalence relations on $\{1, 2, 3, \dots n\}$. Each equivalence class of $\pi(t)$ comprises the labels of all individuals who descend from the progenitor: equivalence classes or blocks correspond to lineages.

**Definition 2.5** ($n$-coalescent)**.** *An $n$-coalescent is the continuous time Markov process $\pi(t)$ on the space of equivalence relations on set $\{1, 2, 3, \dots n\}$ with initial state $\{\{1\}, \{2\}, \{3\}, \dots \{n\}\}$, and transition rates $q_{\xi\eta}$ given by $q_{\xi\eta} = 1$ if $\eta$ is obtained by merging two blocks of $\xi$ and $q_{\xi\eta} = 0$ otherwise.*

Thus the coalescent consists of two elements: the ancestral process and the topology; by the latter we mean that who coalesce with whom.

Since Wright's pioneering work in the 1920s, the W-F has been applied to many cases occurring in nature. Its success is attributed to the fact that even when, with the census population size, W-F does not fit the observed genetic data, it often fits using a different population size, called the effective population size. (One example of this is the case of males and females constituting a diploid population, see Gillespie

(2004) page 49, for example. In §3.4, we will take a similar approach using the results of Nordborg and Krone (2002).) However, cultural transmission of fitness cannot be emulated by simply choosing the right effective population size in W-F. The reason is that effective population size merely laterally rescales the genetic drift. The parents of the alleles in the new generation are still drawn uniformly at random from the previous one; thus we will not see the trans-generational correlation between family sizes which were observed in the cases reviewed above.

## 2.3 Sibert, Austerlitz and Heyer's extension of the Wright-Fisher model

Sibert, Austerlitz and Heyer (2002) extended W-F to capture cultural transmission of fitness: family size influences the offspring numbers. In their model, a parent is still randomly chosen for each individual in the current generation, but not uniformly at random this time.

**Definition 2.6** (Sibert, Austerlitz and Heyer's fixed population sized model)**.** *The chance that the $j$th allele is selected as a parent is weighted by a random function of its family size $S_j$. This weight is $S_j^\alpha X_j$, where $X_j \sim \Gamma(a,1)$ and the parameter $\alpha$ is the measure of "fertility correlation intensity". Here $\Gamma(a,1)$ denotes the Gamma distribution with shape parameter $a > 0$ and scale parameter 1.*

The rationale behind the randomisation of the weights is that inaccuracy in data measurement results in a certain amount of variance: randomisation makes fitting the model to the examined genealogy easier. The authors chose the Gamma distribution because it has a relatively flexible shape, hence providing for different sources of error.

The use of a strictly increasing function of the family size for the parents' weights

does not seem right. It is hard to justify why the fitness of a gene of family size 100 relative to a singleton gene should be $100^{\alpha}$. Indeed, as we pointed out in §1.3, having a large family gradually becomes less advantageous; not to mention that nature imposes a firm bound on humans' fecundity.

Sibert et al. (2002) calls the parameter $\alpha$ the measure of fertility correlation intensity. As we shall see in §4.2, there are two distinct values of the parameter $\alpha$ which yield the same level of correlation because, as $\alpha$ increases, the trans-generational correlation attains its maximum, and then starts to decrease. Indeed, as one would expect, for very large values of $\alpha$, the correlation becomes negative. For this reason, it makes little sense to call $\alpha$ the "measure of fertility correlation intensity".

## 2.4 A tractable model for cultural transmission of fitness

In this subsection we introduce our constant population size model which simplifies Sibert, Austerlitz and Heyer's model and, at the same time, makes it biologically more relevant.

We retain the notation $N$ for total size of our haploid population. We will need to keep track of the number of individuals belonging to different family size classes. We write $p_j^{(N)}(t)$ for the proportion of individuals of family size $j$ in generation $t$. In incorporating cultural transmission of fitness in the Wright-Fisher model, we follow Sibert, Austerlitz and Heyer: we assign parents reproductive weights depending on their family size. The reproduction mechanism of our model is described as follows.

**Definition 2.7** (Wright-Fisher model with CTF)**.** *Consider an haploid population of constant size, $N$. We obtain the new generation of $N$ individuals by selecting a parent for each individual with replacement from the former generation. The chance of picking a given parent of family size $j$ is proportional to the reproductive weight $w_j$.*

Thus the probability that a given individual of family size $j$ is chosen in the parental generation $t$ as mother for an individual in the new generation $t+1$ is $w_j / \sum_{i=1}^{N} w_i N p_i^{(N)}(t)$.

This quantity will often appear in our calculations; in order to simplify the exposition, we introduce the following shorthand notation.

**Definition 2.8** (Infinitesimal relative weights)**.** *We define the infinitesimal relative weight of an individual of family size $j$ in generation $t$, $r_j^{(N)}(t)$, through the equation*

$$\frac{w_j}{\sum_{i=1}^{N} w_i N p_i^{(N)}(t)} = \frac{r_j^{(N)}(t)}{N}.$$

It turns out that the sequence $p = (p_i)$ or equivalently $r = (r_i)$ is the key quantity whose evolution in time fully determines the model's behaviour. Once we know $p$, we know $r$ too, but the use of $r$ greatly simplifies some of our expressions.

**Remark 2.9.** We observe that our model is an extension of W-F. Using the weight sequence $w = (1, 1, 1, \dots)$, all individuals in the parental generation are chosen as parents with equal weight.

**Remark 2.10.** This extension is similar to W-F with selection. Indeed, our model is a case of frequency-dependent selection first observed by Poulton (1884) among certain lepidopterous larvae. Frequency-dependent selection is more widely known to affect virus populations. A rare variant of flu infects more easily, as it is not yet known to the host's immune system. But in our model the reproductive advantage is on certain family sizes instead of allelic types. See §2.5 in Ewens (2004) for a good introduction into frequency-dependent selection.

The difference in the weight of an individual of family size $j$ and of one of size $j+1$ realistically should be small. One possible set of such weights is $(1 + j\alpha)$ for a

daughter of family size $j$. This has the convenient property that, in the case $\alpha = 0$, the individuals in the parental generation are equally likely to be chosen as parents, thus we recover the W-F model. Moreover, this is a good approximation to the exponential weights of Sibert, Austerlitz and Heyer when the parameter $\alpha$ is kept small (the only sensible choice) and $j$ not too big. See Example 3.12 in Part I for an approximation of the family size distribution and of the relative weight sequence belonging to this specific weight sequence.

As we pointed out in §2.3, there is no biological or social reason to insist on increasing the weights for very large families. The analysis is made easy by truncating the model in Definition 2.7 as follows. To the first $K$ family sizes we assign weights $(w_j)_{1\le j\le K}$. Larger family sizes all have the weight $w_K$. This is what we mean by informally saying that the sequence $w$ has a constant tail.

Introducing a special variable $p^{(N)}_{K^*}(t)$ to denote the proportion of individuals whose family size is at least $K$, will further simplify the calculations. We extend our $K^*$-notation to limits of series, as well. For any bounded function $f \in D(\mathbf{R})$, let $\sum_{j=1}^{K^*} f(r^{(N)}_j(t))\, p^{(N)}_j(t)$ denote the finite sum whose last term is $f(r^{(N)}_K(t))\, p^{(N)}_{K^*}(t)$. Since the sequence of the infinitesimal relative weights $r$ also has a constant tail, we have

$$\sum_{j=1}^{K^*} f(r^{(N)}_j(t))\, p^{(N)}_j(t) = \sum_{j=1}^{\infty} f(r^{(N)}_j(t))\, p^{(N)}_j(t).$$

The above truncation of the weight sequence is not only imperative biologically, but also an important instrument in our approach for recovering the model's dual when the population size is large. Part I is devoted to the analysis of this model. In the next chapter, we shall see that the random proportions $p^{(N)}_j(t)$ converge to a deterministic limit $p_j(t)$ as $N$ becomes large for any truncated weight sequence. We shall also specify some conditions on the weights $w = (w_j)$ required for convergence

of $p_j(t)$ in time: practically all weightings sufficiently close to 1 (or to each other) are feasible.

In the introduction we gave a detailed overview of some population data: we could fit this extension to the actual data, and in §§3.3 and 3.4 we do give some basic tools for setting the weights $w$ so that observations match the stationary distribution. However, our primary goal is the mathematical analysis of the model, showing that it exhibits the *qualitative features* attributed to CTF in our overview of the phenomenon.

# 3 Part I: An extension of the Wright-Fisher model to cultural transmission of fitness

In this section, we examine the forward-in-time evolution of the family size distribution. We are interested in both the infinite population size limit and the time stationary behaviour. The state of the population process in generation $t$ is determined by the family size proportions $(p_j^{(N)}(t))$ for $1 \leq j < K$ and $p_{K^*}^{(N)}(t)$, the proportion of individuals with family size at least $K$.

**Definition 3.1.** *Let $(\mathcal{F}_t)$ denote the natural filtration of the process $(p^{(N)}(t))$ defined by $\mathcal{F}_t = \sigma(\{p^{(N)}(s) : s \leq t\})$, where $p_i^{(N)}(t)$ is the proportion of individuals of family size $i$ in generation $t$ of a population of size is $N$.*

## 3.1 Convergence in population size

Truncating the family size structure, that is, using weight sequences of constant tail and sending the population size to infinity significantly simplify the forward-in-time dynamics of the process. We obtain the following law of large numbers type limit result.

**Theorem 3.2.** *The proportion of individuals classified by family size $(p^{(N)}(t))$ converges in distribution, as $N \to \infty$. Moreover, the limit process $(p(t))$ is deterministic and satisfies the recursion*

$$p_j(t+1) = \sum_{i=1}^{K^*} r_i(t) p_i(t) \frac{r_i(t)^{j-1}}{(j-1)!} e^{-r_i(t)}, \tag{3.1}$$

*for all $j \geq 1$.*

Our strategy is to show that the variance of $(p^{(N)}(t+1))$, conditional on the

$\sigma$-algebra $\mathcal{F}_t$, vanishes as the population size becomes large. To this end, we use a simple counting argument to find the conditional moments.

*Proof of Theorem 3.2.* We first observe that $Np_j^{(N)}(t+1)$, the number of individuals of family size $j$ in generation $t+1$, is the sum of the number of offspring whose parent has family size $i$ and $j$ children. The sum is over $i$. Hence, we have

$$\begin{aligned} Np_j^{(N)}(t+1) &= j\left(\sum_{l=1}^{Np_1^{(N)}(t)} I_l^{(1)(j)} + ... + \sum_{l=1}^{Np_{K^*}^{(N)}(t)} I_l^{(K^*)(j)}\right) \\ &= j\sum_{i=1}^{K^*}\sum_{l=1}^{Np_i^{(N)}(t)} I_l^{(i)(j)}, \end{aligned}$$

where for $1 \leq i < k$, $I_l^{(i)(j)} = 1$ if the $l^{th}$ individual of family size $i$ has exactly $j$ offspring, and 0 otherwise. Similarly, $I_l^{(K*)(j)} = 1$ if the $l^{th}$ individual of family size at least $K$ has $j$ children and 0 otherwise. For fixed family sizes $i$ and $j$, the indicators $I_l^{(i)(j)}$ are exchangeable, identically distributed but not independent.

If an individual in generation $t$ is of family size $i$, then the number of its offspring has distribution $Bin(N, r_i^{(N)}(t)/N)$, (and the number of offspring of an individual of family size at least $K$ has distribution $Bin(N, r_K^{(N)}(t)/N)$; recall that $r_k^{(N)} = r_K^{(N)}$ for all $k \geq K$). Thus we have

$$\mathbf{P}\{I_1^{(i)(j)} = 1\} = \binom{N}{j}\left(\frac{r_i^{(N)}(t)}{N}\right)^j\left(1 - \frac{r_i^{(N)}(t)}{N}\right)^{N-j},$$

where $\mathbf{P}$ denotes the conditional probability measure on the $\sigma$-algebra $\mathcal{F}_t$. Taking

conditional expectations of both sides of the recursion, we obtain

$$\begin{aligned}
\mathrm{E}\left[p_j^{(N)}(t+1)|\mathcal{F}_t\right] &= \frac{j}{N}\sum_{i=1}^{K^*}\mathrm{E}\left[Np_i^{(N)}(t)I_1^{(i)(j)}|\mathcal{F}_t\right]\text{; ``taking out a known factor''} \\
&= j\sum_{i=1}^{K^*}p_i^{(N)}(t)\mathrm{E}\left[I_1^{(i)(j)}|\mathcal{F}_t\right] \\
&= j\sum_{i=1}^{K^*}p_i^{(N)}(t)\mathbf{P}\{I_1^{(i)(j)}=1\}\text{; Poisson limit} \\
&= j\sum_{i=1}^{K^*}p_i^{(N)}(t)\left(\frac{r_i^{(N)}(t)^j}{j!}e^{-r_i^{(N)}(t)}+o\left(1\right)\right); \\
&\qquad \text{as the sum is finite, and } p_{K^*,N}(t)\leq 1 \\
&= \sum_{i=1}^{K^*}r_i^{(N)}(t)p_i^{(N)}(t)\frac{r_i^{(N)}(t)^{j-1}}{(j-1)!}e^{-r_i^{(N)}(t)}+o\left(1\right).
\end{aligned} \tag{3.2}$$

In order to find the second moment of the $p_j^{(N)}(t+1)$ conditional on the past, we will need the probability $\mathbf{P}\{I_1^{(i)(j)}I_2^{(k)(j)}=1\}$, that is, the chance that a pair of individuals of family sizes $i$ and $k$, respectively, each have $j$ offspring. To calculate this, we rewrite the event $\{I_1^{(i)(j)}I_2^{(k)(j)}=1\}$ as follows.

1. Select $j$ offspring for the first member of the pair (the individual of family size $i$).

2. Exclude the parent and her $j$ children from the population.

3. Select $j$ offspring for the second member of the pair (the individual of family size $k$) from the remaining population.

We observe that the removal of the first parent slightly alters the relative weight of the second parent: her relative weight becomes $w_k/(w_1Np_1^{(N)}+\cdots+w_i(Np_i^{(N)}-$

$1) + \cdots + w_{K^*} N p_{K^*}^{(N)}) =: r_{ki}^{(N)}/N$. The selections are binomial and by the exclusion step they have been made independent of each other. Using the notation $r_{ki}^{(N)}$ for the altered infinitesimal relative weight of the second parent, we therefore obtain

$$\mathbf{P}\{I_1^{(i)(j)} I_2^{(k)(j)} = 1\} = \binom{N}{j} \left(\frac{r_i^{(N)}(t)}{N}\right)^j \left(1 - \frac{r_i^{(N)}(t)}{N}\right)^{N-j} \cdot \\ \cdot \binom{N-j}{j} \left(\frac{r_{ki}^{(N)}(t)}{N-1}\right)^j \left(1 - \frac{r_{ki}^{(N)}(t)}{N-1}\right)^{N-2j}.$$

Next, we show that the quantity $N(r_k^{(N)} - r_{ki}^{(N)})$ vanishes as $N$ becomes large. Note that $w_i/N \to 0$, as $N \to \infty$, because the weights are assumed not to depend on the population size. Then, using algebra of limits, we obtain

$$N(r_k^{(N)} - r_{ki}^{(N)}) = w_k \left( \frac{1}{\sum_{j=1}^{K^*} w_j p_j^{(N)}} - \frac{1}{w_i \frac{N p_i^{(N)} - 1}{N} + \sum_{j=1, j\neq i}^{K^*} w_j p^{(N)_j}} \right) \\ \to w_k \left( \frac{1}{\sum_{j=1}^{K^*} w_j p_j^{(N)}} - \frac{1}{\sum_{j=1}^{K^*} w_j p_j^{(N)}} \right) = 0,$$

as $N$ becomes large. Thus we have $r_{ki}^{(N)} = r_k^{(N)} + o\,(1/N)$. Using this, we obtain that

$$\mathbf{P}\{I_1^{(i)(j)} I_2^{(k)(j)} = 1\} = \mathbf{P}\{I_1^{(i)(j)} = 1\} \mathbf{P}\{I_2^{(k)(j)} = 1\} + o\,(1)\,. \tag{3.3}$$

We are now ready to find the second moment of $p_j^{(N)}(t+1)$ conditional on the past, $\mathcal{F}_t$. As before, we express the number of individuals of family size $j$ in generation $t+1$ as a sum of indicators.

$$\mathrm{E}\left[p_j^{(N)}(t+1)^2|\mathcal{F}_t\right] = \frac{j^2}{N^2}\mathrm{E}\left[\left(\sum_{l=1}^{Np_1^{(N)}(t)} I_l^{(1)(j)} + ... + \sum_{l=1}^{Np_{K^*}^{(N)}(t)} I_l^{(K^*)(j)}\right)^2 \middle| \mathcal{F}_t\right];$$

expanding the square, and "taking out what is known"

$$\begin{aligned}
&= \frac{j^2}{N^2}\Bigg(\sum_{i=1}^{K^*} Np_i^{(N)}(t)\mathbf{P}\{I_1^{(i)(j)} = 1\} \\
&\quad + \sum_{i=1}^{K^*} Np_i^{(N)}(t)(Np_i(t)^{(N)} - 1)\mathbf{P}\{I_1^{(i)(j)}I_2^{(i)(j)} = 1\} \\
&\quad + 2\sum_{i<k}^{K^*} Np_i^{(N)}(t)Np_k^{(N)}(t)\mathbf{P}\{I_1^{(i)(j)}I_1^{(k)(j)} = 1\}\Bigg) \\
&= j^2\Bigg(O\left(\frac{1}{N}\right) + \sum_{i=1}^{K^*} p_i^{(N)}(t)^2\mathbf{P}\{I_1^{(i)(j)}I_2^{(i)(j)} = 1\} + O\left(\frac{1}{N}\right) + \\
&\quad + 2\sum_{i<k}^{K^*} p_i^{(N)}(t)p_k^{(N)}(t)\mathbf{P}\{I_1^{(i)(j)}I_1^{(k)(j)} = 1\}\Bigg).
\end{aligned}$$

Taking limits of both sides, as $N$ becomes large, and substituting Equation (3.3) in, we obtain that

$$\begin{aligned}
\lim_{N\to\infty} \mathrm{E}\left[p_j^{(N)}(t+1)^2|\mathcal{F}_t\right] &= \lim_{N\to\infty}\left[j^2\left(\sum_{i=1}^{K^*} p_i^{(N)}(t)\mathbf{P}\{I_1^{(i)(j)} = 1\}\right)^2 + O\left(\frac{1}{N}\right)\right] \\
&= j^2\left(\sum_{i=1}^{K^*} r_i(t)p_i(t)\frac{r_i(t)^{j-1}}{(j-1)!}e^{-r_i(t)}\right)^2,
\end{aligned} \tag{3.4}$$

where $p_i(t) = \lim_{N\to\infty} p_i^{(N)}(t)$, and $r_i(t) = \lim_{N\to\infty} r_i^{(N)}(t)$. Equations (3.2) and (3.4) together therefore show that the variance of the proportion of individuals of family size $j$, $p_j^{(N)}(t+1)$, vanishes as the population size becomes large:

$$\lim_{N\to\infty} \operatorname{var}[p_j^{(N)}(t+1)|\mathcal{F}_t] = \lim_{N\to\infty} \mathrm{E}\left[p_j^{(N)}(t+1)^2|\mathcal{F}_t\right] - (\lim_{N\to\infty} \mathrm{E}\left[p_j^{(N)}(t+1)|\mathcal{F}_t\right])^2$$
$$= 0.$$

Therefore the sequence $p_j^{(N)}(t+1)$ converges in distribution to a deterministic sequence (conditional on $\mathcal{F}_t$), as $N \to \infty$. We do not have convergence in $L^2$ because the sample space is different for each value of the population size $N$; nevertheless knowing only the distribution of $p(t) := \lim_{N\to\infty} p^{(N)}(t)$ will suffice for our calculations. In particular, by Equation (3.2), we have the following recursion for the sequence $(p(t))$ in the infinite population size limit.

$$p_j(t+1) = \sum_{i=1}^{K^*} r_i(t) p_i(t) \frac{r_i(t)^{j-1}}{(j-1)!} e^{-r_i(t)}$$

for all $j, t \geq 1$. This completes the proof. □

In §3.4, studying the coalescent dual, we will need the distance between $p_j^{(N)}(t)$ and $p_j(t)$ be not too large. Accordingly, we close the subsection with an approximation result.

**Definition 3.3** (The process $\bar{p}(t)$)**.** *We define the $\mathcal{F}_t$ adapted stochastic process $\bar{p}(t)$ by $\bar{p}_j(t) = N(p_j^{(N)}(t) - p_j(t))$ for all $1 \leq j \leq N$, where $p_j^{(N)}(t)$ denotes the proportion of the $N$ individuals whose family size is $j$ in generation $t$, and $p_j(t) = \lim_{N\to\infty} p_j^{(N)}(t)$.*

**Proposition 3.4.** *For each $t \in \mathbf{N}$, we have the expansion*

$$\begin{aligned} r_j^{(N)}(t) &= r_j(t) - r_j(t)\frac{\sum_{i=1}^{K^*} w_i \bar{p}_i(t)}{\sum_{i=1}^{K^*} w_i p_i(t)}\frac{1}{N} + O(N^{-2}) \\ &= r_j(t) + \frac{\bar{r}_j(t)}{N} + O(N^{-2}), \end{aligned}$$

*denoting the coefficient of $1/N$ by $\bar{r}$, the process $\bar{p}(t)$ satisfies*

$$\begin{aligned} \mathrm{E}\left[\bar{p}_j(t+1)|\mathcal{F}_t\right] = &\sum_{i=1}^{K^*}\left[p_i(t)\left(-\frac{r_i(t)^2}{2} + jr_i(t) - \frac{j(j-1)}{2}\right) + \bar{p}_i(t)\right]\frac{r_i(t)^j}{(j-1)!}e^{-r_i(t)} + \\ &+ \sum_{i=1}^{K^*} p_i(t)\bar{r}_i(t)\frac{jr_i(t)^{j-1} - r_i(t)^j}{(j-1)!}e^{-r_i(t)} \end{aligned}$$

*for all $j$ and $t \in \mathbf{N}^+$.*

*Proof.* To prove the first claim, regarding the expansion of the relative weight $r_j^{(N)}$, we calculate

$$\begin{aligned} r_j^{(N)}(t) &= \frac{w_j}{\sum_{i=1}^{K^*} w_i p_i^{(N)}(t)} && \text{definition of the relative weights } r_j^{(N)}; \\ &= \frac{w_j}{\sum_{i=1}^{K^*} w_i \left(p_i(t) + \frac{\bar{p}_j(t)}{N}\right)} && \text{definition of the process } \bar{p}_j; \\ &= \frac{w_j}{\sum_{i=1}^{K^*} w_i p_i(t) + \frac{1}{N}\sum_{i=1}^{K^*} w_i \bar{p}_j(t)}; && \text{geometric expansion} \\ &= \frac{w_j}{\sum_{i=1}^{K^*} w_i p_i(t)}\left(1 - \frac{1}{N}\frac{\sum_{i=1}^{K^*} w_i \bar{p}_i(t)}{\sum_{i=1}^{K^*} w_i p_i(t)} + O(N^{-2})\right); && \text{defn. of } r_j(t) \\ &= r_j(t) - r_j(t)\frac{\sum_{i=1}^{K^*} w_i \bar{p}_i(t)}{\sum_{i=1}^{K^*} w_i p_i(t)}\frac{1}{N} + O(N^{-2}). \end{aligned}$$

To verify the second claim, we recall that in the proof of Theorem 3.2 we calculated that

$$\mathrm{E}\left[p_j^{(N)}(t+1)|\mathcal{F}_t\right] = j\sum_{i=1}^{K^*} p_i^{(N)}(t)\mathbf{P}\{I_1^{(i)(j)}=1\}, \tag{3.5}$$

where we are using the indicator $I_l^{(i)(j)}$ for the event that the $l$th individual of family size $i$ in generation $t$ gives birth to $j$ daughters. From the proof of Theorem 3.2, using power series expansions, we obtain that

$$\begin{aligned}
\mathbf{P}\{I_1^{(i)(j)}=1\} &= \binom{N}{j}\left(\frac{r_i^{(N)}(t)}{N}\right)^j\left(1-\frac{r_i^{(N)}(t)}{N}\right)^{N-j}\\
&= \binom{N}{j}\left(\frac{r_i^{(N)}}{N}\right)^j \exp\left((N-j)\log\left(1-\frac{r_i^{(N)}}{N}\right)\right)\\
&= \binom{N}{j}\left(\frac{r_i^{(N)}}{N}\right)^j \exp\left(-(N-j)\frac{r_i^{(N)}}{N}-(N-j)\frac{{r_i^{(N)}}^2}{2N^2}-\dots\right)\\
&= \binom{N}{j}\left(\frac{r_i^{(N)}}{N}\right)^j e^{-r_i^{(N)}}\exp\left(\frac{jr_i^{(N)}-{r_i^{(N)}}^2/2}{N}+O(N^{-2})\right)\\
&= \binom{N}{j}\left(\frac{r_i^{(N)}}{N}\right)^j e^{-r_i^{(N)}}\left(1+\frac{jr_i^{(N)}-{r_i^{(N)}}^2/2}{N}+O(N^{-2})\right)\\
&= \frac{1\left(1-\frac{1}{N}\right)\dots\left(1-\frac{j-1}{N}\right)}{j!}{r_i^{(N)}}^j e^{-r_i^{(N)}}\left(1+\frac{jr_i^{(N)}-{r_i^{(N)}}^2/2}{N}+O(N^{-2})\right)\\
&= \left(1-\sum_{k=1}^{j-1}\frac{k}{N}+O(N^{-2})\right)\frac{{r_i^{(N)}}^j}{j!}e^{-r_i^{(N)}}\left(1+\frac{jr_i^{(N)}-{r_i^{(N)}}^2/2}{N}+O(N^{-2})\right)\\
&= \frac{{r_i^{(N)}}^j}{j!}e^{-r_i^{(N)}}\left[1-\left(\frac{{r_i^{(N)}}^2}{2}-jr_i^{(N)}+\frac{j(j-1)}{2}\right)\frac{1}{N}+O(N^{-2})\right],
\end{aligned}$$

where we have suppressed the dependence on the generation number $t$. The reason that we are able to use the Maclaurin series expansion for the function $\log(1-x)$ above is that the population size $N$ is assumed to be large and we work with finitely many different weights $w_i$, making the relative weights $r_i^{(N)}$ bounded, so

$|r_i^{(N)}/N| < 1$. Then we rewrite the factors in this product as

$$1 - \left( \frac{{r_i^{(N)}}^2}{2} - j r_i^{(N)} + \frac{j(j-1)}{2} \right) \frac{1}{N} = 1 - \left( \frac{{r_i}^2}{2} - j r_i + \frac{j(j-1)}{2} \right) \frac{1}{N} + O(N^{-2});$$

and

$$\begin{aligned} {r_i^{(N)}}^j e^{-r_i^{(N)}} &= \left( r_i + \frac{\bar{r}_i}{N} + O(N^{-2}) \right)^j e^{-r_i} \exp\left( -\frac{\bar{r}_i}{N} + O(N^{-2}) \right) \\ &= \left( r_i^j + j \frac{\bar{r}_i r_i^{j-1}}{N} + O(N^{-2}) \right) e^{-r_i} \left( 1 - \frac{\bar{r}_i}{N} + O(N^{-2}) \right) \\ &= r_i^j e^{-r_i} + \frac{(j r_i^{j-1} - r_i^j)\bar{r}_i}{N} e^{-r_i} + O(N^{-2}). \end{aligned}$$

Substituting now back into Equation (3.5) and matching the coefficients of $1/N$ on the two sides, we obtain the expression for $\mathrm{E}\,[\bar{p}_j^{(N)}(t+1)|\mathcal{F}_t]$ in the statement of the proposition. This completes the proof. □

The above deterministic limit results are still with discrete generations, whereas fixed size genealogy has a scaling limit ancestral process. It is possible however that there is a stochastic diffusion process limit for $p^{(N)}(Nt)$, the family size proportions run on the time scale of the population size, if the weights $w_i$ have the form $1+\psi_i/N$ for some sequence $\psi$ uniformly bounded in $N$. This is by analogy with the usual weak selection diffusion limit in population genetics. See §§ 2.5 and 5.3 in Ewens (2004).

In the next section we find some conditions under which the family size has a time-stationary distribution.

## 3.2 Convergence in time

Whether the infinite population limit of the family size proportions $p(t)$ converge in time depends on the weight sequence. In §3.4 we study the coalescent process associated with the model. In order to keep the analysis there simple, we will assume that family size has a stationary distribution. What we see in practice is that for weight sequences close to $w = (1, 1, 1, \dots)$ or, in other words, in the W-F case $p(t)$ converges to a probability distribution as $t \to \infty$, while if we assign very large weights to some family size classes relative to the rest, then $p(t)$ will have a limit-cycle. This phenomenon is further explored in Example 3.9, but we will focus here on finding some sufficient conditions for convergence.

In the previous section we obtained a recursion for the family size class proportions $p(t)$: for all $t \in \mathbf{N}$,

$$p_j(t+1) = \sum_{i=1}^{K^*} r_i(t) p_i(t) \frac{r_i(t)^{j-1}}{(j-1)!} e^{-r_i(t)}.$$

The proportions $(p_j(t))$ will converge as $t$ becomes large if we can show that the function $f : [0,1]^{K-1} \to [0,1]^{K-1}$, defined by $f((p_j(t))) = (p_j(t+1))$ is a contraction. This means that for a predetermined metric $d$ on the set $[0,1]^{K-1}$ there is a constant $s \in (0,1)$ such that

$$d(f(x), f(y)) \leq s\, d(x, y)$$

for all $x, y \in [0,1]^{K-1}$. In general, the function $f$ is not a contraction, but below we find some conditions on the weight sequences under which, $f$ is a contraction mapping.

We illustrate our strategy by first considering the easy case K = 2. Now the map

$f : [0,1] \to [0,1]$ is given by

$$f(x) = xr_1(x)e^{-r_1(x)} + (1-x)r_2(x)e^{-r_2(x)},$$

for any $x \in [0,1]$. The argument $x$ in the notation $r_j(x)$ stresses that the infinitesimal relative weight of daughters of family size $j$ does depend on $x$, the proportion of singletons (individuals of family size one) in the population. Let $x, y \in [0,1]$. Then using the triangle inequality, we obtain that

$$\begin{aligned}
|f(x) - f(y)| &= |xr_1(x)e^{r_1(x)} + (1-x)r_2(x)e^{r_2(x)} - yr_1(y)e^{r_1(y)} - (1-y)r_2(y)e^{r_2(y)}| \\
&= |(xr_1(x)e^{r_1(x)} - yr_1(x)e^{r_1(x)}) + (yr_1(x)e^{r_1(x)} - yr_1(y)e^{r_1(y)}) + \\
&\quad + ((1-x)r_2(x)e^{r_2(x)} - (1-y)r_2(x)e^{r_2(x)}) + \\
&\quad + ((1-y)r_2(x)e^{r_2(x)} - (1-y)r_2(y)e^{r_2(y)})| \\
&\leq |x-y||r_1(x)e^{r_1(x)} - r_2(x)e^{r_2(x)}| + \\
&\quad + y|r_1(x)e^{r_1(x)} - r_1(y)e^{r_1(y)}| + (1-y)|r_2(x)e^{r_2(x)} - r_2(y)e^{r_2(y)}| \\
&\leq |x-y||r_1(x)e^{r_1(x)} - r_2(x)e^{r_2(x)}| + \\
&\quad + \max_{j=1,2}\{|r_j(x)e^{r_j(x)} - r_j(x)e^{r_j(x)}|\}(y + 1 - y) \\
&\leq s|x-y|,
\end{aligned}$$

provided $\max_{j=1,2}\{|r_j(x)e^{r_j(x)} - r_j(y)e^{r_j(y)}|\} \leq q|x-y|$ and $|r_1(x)e^{r_1(x)} - r_2(x)e^{r_2(x)}| \leq s - q$, for some $q$ and $s \in \mathbf{R}^+$ such that $q < s < 1$.

The above strategy works in the general case, $K \geq 2$, too. We define the map $r : [0,1]^{K-1} \to [0,1]^{K-1}$ by, for all $x \in [0,1]^{K-1}$,

$$r_i(x) = \frac{w_i}{\sum_{j=1}^{K} w_j x_j}$$

with $x_K = 1 - \sum_{j=1}^{K-1} x_j$. Thus the function $r$ is fully determined by the set of weights $\{w_1, \dots w_K\}$. Choosing different metrics on the $(K-1)$-dimensional space gives different sufficient conditions on the $r_j$'s. Among the many metrics that the space admits, the ones generated by the $l^\infty$- and $l^1$-norms seem to be the easiest to manipulate.

**Theorem 3.5.** *The family size distribution $(p_j(t))$ for large populations converge to a time stationary distribution at geometric rate as time passes if there exist a constant $a \in [0,1)$ and a function $b : [0,1]^{K-1} \to [0,1)$ such that $s := a + \sup_{x\in[0,1]^{K-1}} b(x) < 1$ and either of the following two conditions is satisfied.*

1. $\max_{\substack{1\le j\le K-1\\ 1\le m\le K}} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right| \le a\|x-y\|_\infty$ *and*
   $\max_{1\le j\le K-1} \sum_{m=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_K(x)^j}{(j-1)!} e^{-r_K(x)} \right| \le b(x)$, *or*

2. $\max_{1\le m\le K} \sum_{j=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right| \le a\|x-y\|_1$ *and*
   $\max_{1\le m\le K} \sum_{j=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_K(x)^j}{(j-1)!} e^{-r_K(x)} \right| \le b(x)$.

See the proof in Appendix A.

Returning to the case $K = 2$, that is, if we distinguish between only two groups: singletons and individuals only then we can find an explicit condition for convergence of the family size distributions $(p(t))$. Now the function $f$ is differentiable and has only one variable, hence we can decide whether it is a contraction by studying its derivative. In order to simplify the calculation, we may assume without loss of generality that the second weight $w_2$ equals 1, (as any scalar multiple of the weight sequence $w$ has the same effect.) Thus we need to study two cases only: $w_1 < 1$ or $w_1 > 1$, the case $w_1 = 1$ being trivial.

**Proposition 3.6.** *In the case $K = 2$ assuming that the weights $w_1$ and $w_2$ are normalised such that $w_2 = 1$, the sequence of family size distributions $(p(t))$ converges as $t \to \infty$, if*

$$\frac{|w_1 - 1|}{\min(w_1, 1)} M_1(w_1) + \frac{|w_1 - 1|^2}{\min(w_1, 1)^2} M_1(w_1) + \\ + \frac{|w_1 - 1|^2}{\min(w_1, 1)^2} M_2(w_1) + \frac{|w_1 - 1|}{\min(w_1, 1)} M_3(w_1) < 1,$$

*for some functions $M_1$, $M_2$ and $M_3$ specified below. If $0 < w_1 < w_2 = 1$, then*

$$M_1(w_1) = \frac{\max\{\cosh w_1^{-1} - w_1 \cosh 1,\ \sinh w_1^{-1} - w_1 \sinh 1\}}{1 - w_1} - e;$$
$$M_2(w_1) = \frac{1}{1 - w_1} \max\{w_1^{-1} \cosh w_1^{-1} - w_1 \cosh 1 - \sinh 1, \\ w_1^{-1} \sinh w_1^{-1} - w_1 \sinh 1 - \cosh 1\} - e;$$
$$M_3(w_1) = \frac{1 - w_1}{w_1^2} e^{-1}.$$

*In the case when $w_1 \geq w_2 = 1$,*

$$M_1(w_1) = \max\left\{ \frac{w_1 \cosh w_1 - \cosh 1}{w_1 - 1} - \sinh w_1^{-1}, \right. \\ \left. \frac{w_1 \sinh w_1 - \sinh 1}{w_1 - 1} - \cosh w_1^{-1} \right\} - w_1^{-1} \sinh w_1^{-1};$$
$$M_2(w_1) = \max\left\{ \frac{w_1^2 \cosh w_1 - \cosh 1}{w_1 - 1} w_1^{-1} \cosh w_1^{-1} - 2 \sinh w_1^{-1}, \right. \\ \left. \frac{w_1^2 \sinh w_1 - \sinh 1}{w_1 - 1} - 2 w_1^{-1} \cosh w_1^{-1} - w_1^{-2} \sinh w_1^{-1} \right\};\ \textit{and}$$
$$M_3(w_1) = (w_1 - 1) e^{-w_1^{-1}}.$$

*Proof.* With the notation above, we prove that the function $f$ is a contraction by using the Mean Value Theorem for derivatives and the fact that $|f'(x)| < 1$, for all

$x \in (0,1)$.

First we calculate the derivative of the relative weights $r_1$ and $r_2$ with respect to $x$. For $i = 1, 2$, we obtain

$$r_i'(x) = \left(\frac{w_i}{w_1 x + w_2(1-x)}\right)' = -\frac{(w_1 - w_2)w_i}{(w_1 x + w_2(1-x))^2} = (r_2(x) - r_1(x))r_i(x).$$

For clarity we suppress the dependence of the relative weights $r$ on $x$ in our notation. Now the Chain Rule gives us that

$$\begin{aligned} f'(x) &= r_1 e^{-r_1} + x(r_1 e^{-r_1})' - r_2 e^{-r_2} + (1-x)(r_2 e^{-r_2})' \\ &= r_1 e^{-r_1} - r_2 e^{-r_2} + x r_1' e^{-r_1} - x r_1 r_1' e^{-r_1} + (1-x) r_2' e^{-r_2} - (1-x) r_2 r_2' e^{-r_2}. \end{aligned}$$

Substituting in $r'$ and using the triangle inequality, we obtain that

$$\begin{aligned} |f'(x)| \leq\ & |r_1 e^{-r_1} - r_2 e^{-r_2}| + \\ & + |x||r_1 - r_2||r_1 e^{-r_1} - r_2 e^{-r_2}| + \\ & + |x||r_1 - r_2||r_1^2 e^{-r_1} - r_2^2 e^{-r_2}| + \\ & + |r_1 - r_2| r_2 e^{-r_2} |1 - r_2|. \end{aligned} \tag{3.6}$$

Next, we estimate the four terms separately. For the first term, we use the power

series expansion of the exponential function.

$$\begin{aligned} r_1 e^{-r_1} - r_2 e^{-r_2} &= \sum_{j=0}^{\infty} \frac{r_1(-r_1)^j - r_2(-r_2)^j}{j!}; \text{ using the identity } \text{“}a^n - b^n\text{”} \\ &= (r_1 - r_2) \sum_{j=0}^{\infty} \frac{(-r_1)^j + (-r_1)^{j-1}(-r_2) + \cdots + (-r_2)^j}{j!} \\ &= (r_1 - r_2) \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \sum_{k=0}^{j} r_1^{j-k} r_2^k. \end{aligned}$$

We observe that all scalar multiples of a weight sequence have the same effect. Thus, we are free to choose the weight $w_2$ to be one. Furthermore, the quantity $w_1 x + w_2(1-x)$ is a convex combination of the weights $w_1$ and $w_2 = 1$, and hence we have both a lower and an upper bound for the inner sum in terms of $w_1$ alone.

$$\frac{j+1}{\max(w_1, 1)} \leq \sum_{k=0}^{j} r_1^k r_2^{j-k} \leq \sum_{k=0}^{j} \frac{w_1^k}{\min(w_1, 1)}.$$

Thus, if $w_1 > w_2 = 1$, then $1 \leq x w_1 + (1-x) \cdot 1 \leq w_1$ and so

$$\begin{aligned} r_1 e^{-r_1} - r_2 e^{-r_2} &= (r_1 - r_2) \left( \sum_{j=0}^{\infty} \frac{(-1)^{2j}}{(2j)!} \sum_{k=0}^{2j} r_1^k r_2^{j-k} + \sum_{j=0}^{\infty} \frac{(-1)^{2j+1}}{(2j+1)!} \sum_{k=0}^{2j+1} r_1^k r_2^{j-k} \right) \\ &\leq (r_1 - r_2) \left( \sum_{j=0}^{\infty} \frac{1}{(2j)!} \sum_{k=0}^{2j} w_1^k - \sum_{j=0}^{\infty} \frac{2j+2}{(2j+1)!} w_1^{-(2j+1)} \right). \end{aligned}$$

Rewriting the series in terms of hyperbolic functions, we obtain that

$$\begin{aligned} r_1 e^{-r_1} - r_2 e^{-r_2} &\leq (r_1 - r_2) \left( \sum_{j=0}^{\infty} \frac{1}{(2j)!} \frac{1 - w_1^{2j+1}}{1 - w_1} - \sum_{j=0}^{\infty} \frac{1}{(2j+1)!} \frac{2j+2}{w_1^{2j+1}} \right) \\ &= (r_1 - r_2) \left( \frac{1}{1 - w_1} \left( \sum_{j=0}^{\infty} \frac{1}{(2j)!} - w \sum_{j=0}^{\infty} \frac{w_1^{2j}}{(2j)!} \right) - \sum_{j=0}^{\infty} \frac{2j+2}{(2j+1)!} w_1^{-(2j+1)} \right) \\ &= (r_1 - r_2) \left( \frac{\cosh 1 - w_1 \cosh w_1}{1 - w_1} - (x \sinh x)'|_{x = w_1^{-1}} \right) \\ &= (r_1 - r_2) \left( \frac{\cosh 1 - w_1 \cosh w_1}{1 - w_1} - \sinh w_1^{-1} - \frac{\sinh w_1^{-1}}{w_1} \right). \end{aligned}$$

Likewise we obtain that

$$\begin{aligned} -(r_1 e^{-r_1} - r_2 e^{-r_2}) &\leq (r_2 - r_1) \left( \sum_{j=0}^{\infty} \frac{1}{(2j+1)!} \sum_{k=0}^{2j+1} w_1^k - \sum_{j=0}^{\infty} \frac{1}{(2j)!} \frac{2j+1}{w_1^{2j}} \right) \\ &= (r_2 - r_1) \left( \frac{\sinh 1 - w_1 \sinh w_1}{1 - w_1} - \cosh w_1^{-1} - \frac{\sinh w_1^{-1}}{w_1} \right). \end{aligned}$$

Hence we find an upper bound, $M_1(w_1)$, for the first term

$$\begin{aligned} &|r_1 e^{-r_1} - r_2 e^{-r_2}| \leq |r_1 - r_2| \cdot \\ &\cdot \max \left\{ \sum_{j=0}^{\infty} \frac{1}{(2j)!} \sum_{k=0}^{2j} w_1^k - \sum_{j=0}^{\infty} \frac{2j+2}{(2j+1)!} w_1^{-(2j+1)}, \sum_{j=0}^{\infty} \frac{1}{(2j+1)!} \sum_{k=0}^{2j+1} w_1^k - \sum_{j=0}^{\infty} \frac{1}{(2j)!} \frac{2j+1}{w_1^{2j}} \right\} \\ &= |r_1 - r_2| \max \left\{ \frac{\cosh 1 - w_1 \cosh w_1}{1 - w_1} - \sinh w_1^{-1} - \frac{\sinh w_1^{-1}}{w_1}, \right. \\ &\qquad \left. \frac{\sinh 1 - w_1 \sinh w_1}{1 - w_1} - \cosh w_1^{-1} - \frac{\sinh w_1^{-1}}{w_1} \right\} \\ &:= |r_1 - r_2| M_1(w_1). \end{aligned}$$

If $0 < w_1 < w_2 = 1$, then we find that

$$
\begin{aligned}
&|r_1 e^{-r_1} - r_2 e^{-r_2}| \le |r_1 - r_2| \cdot \\
&\cdot \max\left\{ \sum_{j=0}^{\infty} \frac{w_1^{-2j}}{(2j)!} \sum_{k=0}^{2j} w_1^k - \sum_{j=0}^{\infty} \frac{2j+2}{(2j+1)!}, \sum_{j=0}^{\infty} \frac{w_1^{-(2j+1)}}{(2j+1)!} \sum_{k=0}^{2j+1} w_1^k - \sum_{j=0}^{\infty} \frac{2j+1}{(2j)!} \right\} \\
&= |r_1 - r_2| \ \max\left\{ \frac{\cosh w_1^{-1} - w_1 \cosh 1}{1 - w_1} - e, \ \frac{\sinh w_1^{-1} - w_1 \sinh 1}{1 - w_1} - e \right\} \\
&:= |r_1 - r_2| M_1(w_1).
\end{aligned}
$$

The factor $|r_1^2 e^{-r_1} - r_2^2 e^{-r_2}| = |r_1 - r_2||\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \sum_{k=0}^{j+1} r_1^k r_2^{j-k}|$ in the third term is estimated similarly ($M_2$). For the values of the weight $w_1 \ge 1$, we obtain

$$
\begin{aligned}
&|r_1 e^{-r_1} - r_2 e^{-r_2}| \le |r_1 - r_2| \cdot \\
&\cdot \max\left\{ \sum_{j=0}^{\infty} \frac{1}{(2j)!} \sum_{k=0}^{2j+1} w_1^k - \sum_{j=0}^{\infty} \frac{2j+3}{(2j+1)!} w_1^{-(2j+1)}, \sum_{j=0}^{\infty} \frac{1}{(2j+1)!} \sum_{k=0}^{2j+2} w_1^k - \sum_{j=0}^{\infty} \frac{2j+2}{(2j)!} w_1^{-(2j+1)} \right\} \\
&= |r_1 - r_2| \max\left\{ \frac{\cosh 1 - w_1^2 \cosh w_1}{1 - w_1} - \frac{\cosh w_1^{-1}}{w_1} - 2 \sinh w_1^{-1}, \right. \\
&\qquad \left. \frac{\sinh 1 - w_1^2 \sinh w_1}{1 - w_1} - 2\frac{\cosh w_1^{-1}}{w_1} - \frac{\sinh w_1^{-1}}{w_1^2} \right\} \\
&:= |r_1 - r_2| M_2(w_1).
\end{aligned}
$$

For the case $0 < w_1 < 1$, we obtain

$$
\begin{aligned}
&|r_1 e^{-r_1} - r_2 e^{-r_2}| \le |r_1 - r_2| \cdot \\
&\cdot \max\left\{ \sum_{j=0}^{\infty} \frac{w_1^{-(2j+1)}}{(2j)!} \sum_{k=0}^{2j+1} w_1^k - \sum_{j=0}^{\infty} \frac{2j+3}{(2j+1)!}, \sum_{j=0}^{\infty} \frac{w_1^{-(2j+2)}}{(2j+1)!} \sum_{k=0}^{2j+2} w_1^k - \sum_{j=0}^{\infty} \frac{2j+2}{(2j)!} \right\} \\
&= |r_1 - r_2| \max\left\{ \frac{w_1^{-1}\cosh w_1^{-1} - w_1 \cosh 1}{1 - w_1} - e - \sinh 1 \, , \right. \\
&\qquad\qquad \left. \frac{w_1^{-1}\sinh w_1^{-1} - w_1 \sinh 1}{1 - w_1} - e - \cosh 1 \right\} \\
&:= |r_1 - r_2| M_2(w_1).
\end{aligned}
$$

For $w_1 \ge w_2 = 1$, the factor $|r_1 - r_2| r_2 e^{-r_2}$ in the fourth term is estimated by $M_3(w_1)$ as follows. For all $x \in [0,1]$, we have

$$
\begin{aligned}
|r_1 - r_2| r_2 e^{-r_2} &= \frac{|1 - w_1|}{w_1 x + (1 - x)} r_2 e^{-r_2} \\
&= (w_1 - 1) r_2^2 e^{-r_2} \le (w_1 - 1) e^{-\frac{1}{w_1}} \\
&:= M_3(w_1).
\end{aligned}
$$

If $0 < w_1 < w_2 = 1$, then $w_1 \le w_1 x + (1 - x) \le 1$, and hence $1 \le r_2 \le 1/w_1$. Thus, in this case, we obtain that

$$
\begin{aligned}
|r_1 - r_2| r_2 e^{-r_2} = |1 - w_1| r_2^2 e^{-r_2} &\le (1 - w_1) \frac{e^{-1}}{w_1^2} \\
&:= M_3(w_1).
\end{aligned}
$$

It only remains to estimate the factor $|1 - r_2|$ in the fourth term.

$$\begin{aligned} |1 - r_2| &= \frac{|w_1 x + w_2(1-x) - w_2|}{|w_1 x + w_2(1-x)|} = \frac{|w_1 - w_2|}{w_1 x + w_2(1-x)|}|x| \\ &\leq \frac{|w_1 - 1|}{\min(w_1, 1)}. \end{aligned}$$

Substituting the above into Inequality (3.6) completes the proof. □

**Remark 3.7.** The above proposition gives sufficient conditions for convergence of the proportions $p_j(t)$ in the case $K = 2$. Unfortunately $p_j(t)$ seems to converge for all values of $w_1$. To assess the domain for which values $w_1$ in the normalised weight sequence $(w_1, 1, 1, \dots)$, starting with $p_1(0) = 1$, $\lim_{t\to\infty} p(t)$ exists, we used the recursion to compute $p$ up to the first generation $t$ for which $\|p(t) - p(t-1)\|_1 < 10^{-5}$. We found that $t$ strictly monotone increases with $w_1$, for $w_1 \geq 1$. Plotting $(w_1, t)$ on the log-log scale we inspected that

$$\log(t) \approx m \log(w_1) + a$$

for some constants $m \approx 1.0412$ and $a \approx 3.1371$. Hence

$$t \approx \exp(\log(w_1^m) + a) = A w_1^m$$

with $A \approx 23.0370$. When $0 < w_1 < 1$, we reach the stationary distribution in $t = 17$ generations for all values of $w_1 = 10^{-n}$ with $n \geq 3$. We have convergence also at $w_1 = 0$ provided that $p_1(0) < 1$.

**Remark 3.8.** We observe that in the more complicated case $K \geq 3$ the conditions for convergence of $p(t)$ in Theorem 3.5 fail when the infinitesimal relative weights $r_j$'s lie far apart from each other. This, in turn, happens when the weights $w_j$'s do

so, because $r_j$ is proportional to $w_j$. The next example shows that the weights being very far from each other may cause the proportions $p_j(t)$ to have limit-cycles.

**Example 3.9.** Consider the simple case with family size resolution $K = 3$ and weights $(1, w_2, 1, 1, ...)$. Through numerical experiments we see that, for $w_2 \leq 21.77$, the sequence of proportions $(p_j(t))$ converges, while when the weights are sufficiently far apart, for $w_2 \geq 21.78$, $(p_j(t))$ is alternating as $t$ becomes large. Figure 2 shows the time stationary behaviour of $(p_j(t))$ for three different weight sequences: $(1, 1, 1, ...)$, $(1, 20, 1, ...)$, and $(1, 24, 1, ...)$. For the last one, $(p_j(t))$ is alternating, thus $(\lim_{t\to\infty} p_j(2t+1))_{j\geq 1}$ and $(\lim_{t\to\infty} p_j(2t))_{j\geq 1}$ are shown separately in the diagram.

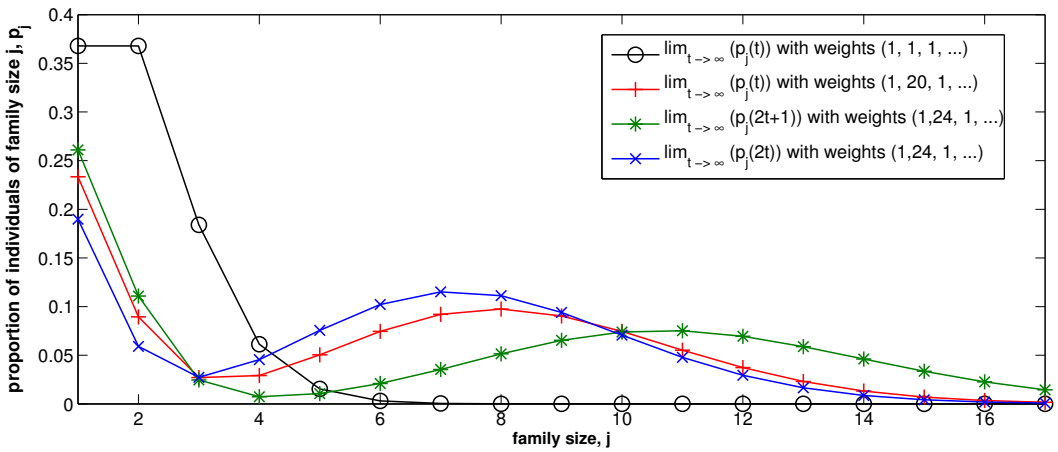



Figure 2: Stationary behaviour of the proportions of genes of family size $j$, for different weight sequences $(w_j)$

The alternating proportions of daughters of a certain family size is a legitimate feature of the model. Should we have many individuals of family size two with a very large propensity to reproduce, we expect that in the next generation their many children (of large family sizes) decreases the number of individuals of family size two because the total population size is constant. In the third generation, however, most of the individuals have equal chance of becoming parents, and hence most of the

offspring are born into small families. Hence the proportion of individuals of family size two will be large again.

**Example 3.10.** It is important to see whether our model can produce similar family size distributions to the ones observed in real human populations where CTF is assumed to be present. In Figure 3, setting the weights $w$ to be $(1/2, 1, 16, 1, 4, 4, 1, 1, ...)$, we attempt to mimic the flat feature of the stationary family size distribution around the 'optimal' family size 5 as seen in the Saguenay-Lac-Saint-Jean region, Québec. Indeed, $p(t)$ becomes time-stationary within a few generations: using the initial condition $p_1(0) = 1$, we obtain that $\|p(25) - p(24)\| < 10^{-5}$. Compare this with the distribution of observed family sizes in Figure 1 in §1.3.

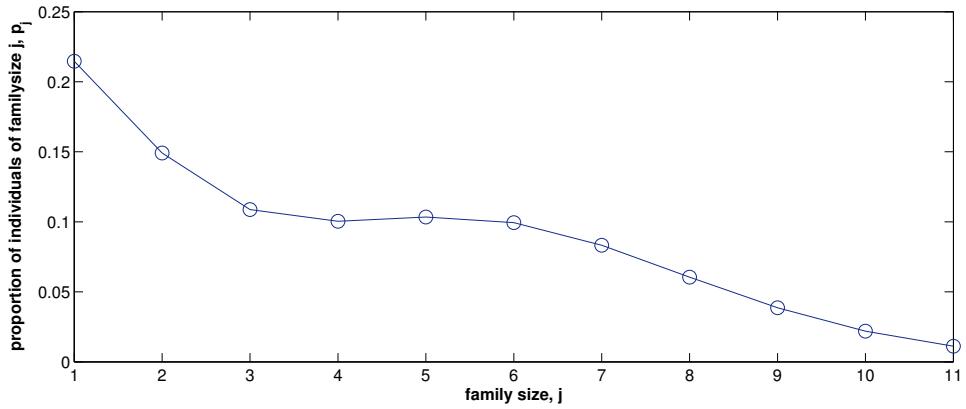



Figure 3: Stationary behaviour of the proportions of genes of family size $j$, for the weight sequence $w = (1/2, 1, 16, 1, 4, 4, 1, 1, ...)$.

We should note here that the family size distribution in itself does not carry enough information that we are able to recover the underlying weight sequence. In the next section, however, we shall see that the family size distribution and the parental family size distribution together do specify $w$ (up to a positive scalar multiple). We conclude this section with a simple estimation of the proportions $p_i(t)$ in the special case when the weights are all close to one.

**Proposition 3.11** (Perturbation of $p(w)$ and $r(w)$ about the Wright-Fisher case)**.** *Suppose that the weights are of the form $w_j = 1 + \epsilon_j$, where $|\epsilon_j| < 1$ is small, for all $j$, $1 \leq j \leq K$. Then the infinitesimal relative weights $r$, and the the proportion of individuals in the different family size classes, $p$ may be approximated as*

$$r_j(t) = 1 + \epsilon_j - \sum_{i=1}^{\infty} \frac{\epsilon_i}{(i-1)!} e^{-1} + O(\|\epsilon\|_\infty^2), \text{ and}$$
$$p_j(t) = \frac{e^{-1}}{(j-1)!} + O_j(\|\epsilon\|_\infty^2),$$

*where we used the notation $O_j$ to signify that the error term is summable over $j$ with the sum still $O(\|\epsilon\|_\infty^2)$.*

*Indeed, for small values of $\epsilon$, $p(t)$ converges in $t$ and it attains its limit within an error $O(\|\epsilon\|_\infty^2)$ in the second generation, just as it does in the Wright-Fisher case.*

The result tells us that small deviations from the Wright-Fisher case do not affect the family size. Thus, in Japan or in the United States of America, for example, we should expect to find that the family size of a randomly picked individual has a distribution very close to $Pois(1) + 1$. Interestingly, although the family size distribution $p$ remains largely unaffected, we will see in §3.4 that the speed at which ancestral lines coalesce (time measured in units of the population size) *is affected* by the sequence $\epsilon$.

*Proof of Proposition 3.11.* We use the geometric series expansion $\sum_{k=0}^{\infty} x^k = 1/(1-$

$x$) to estimate $r$. For all $i, t \geq 1$, we have

$$\begin{aligned}
r_i(t) &= \frac{w_i}{\sum_{j=1}^{\infty} w_j p_j(t)};\ \text{assumption} \\
&= \frac{1+\epsilon_i}{\sum_{j=1}^{\infty}(1+\epsilon_j)p_j(t)};\ \text{as}\ \sum_{j=1}^{\infty} p_j = 1 \\
&= \frac{1+\epsilon_i}{1+\sum_{j=1}^{\infty}\epsilon_j p_j(t)};\ \text{geometric expansion as}\ \sum_{j=1}^{\infty}|\epsilon_j|p_j < \sum_{j=1}^{\infty} p_j = 1 \\
&= (1+\epsilon_i)\sum_{k=0}^{\infty}\left(-\sum_{j=1}^{\infty}\epsilon_j p_j(t)\right)^k \\
&= (1+\epsilon_i)(1-\sum_{j=1}^{\infty}\epsilon_j p_j(t) + O(\|\epsilon\|_\infty^2));\ \text{as}\ p_j(t) \in [0,1] \\
&= 1+\epsilon_i - \sum_{j=1}^{\infty}\epsilon_j p_j(t) + O(\|\epsilon\|_\infty^2).
\end{aligned}$$

To justify the order of the error terms in the last two lines, we note that

$$\left(\sum_{j=1}^{\infty}\epsilon_j p_j(t)\right)^2 \leq \left(\sup_j|\epsilon_j|\sum_{j=1}^{\infty} p_j(t)\right)^2 = \|\epsilon\|_\infty^2 \cdot 1$$

and

$$|\epsilon_i\sum_{j=1}^{\infty}\epsilon_j p_j(t)| \leq |\epsilon_i|\sum_{j=1}^{\infty}|\epsilon_j|p_j(t) \leq |\epsilon_i|\sup_j|\epsilon_j|\sum_{j=1}^{\infty} p_j(t) \leq \|\epsilon\|_\infty^2 \cdot 1.$$

We use this approximation of $r$ to estimate $p$. By Recursion (3.1), for all $j$ and

$t \geq 1$, we have

$$
\begin{aligned}
p_j(t+1) &= \sum_{i=1}^{K^*} p_i(t) \frac{r_i(t)^j}{(j-1)!} e^{-r_i(t)}; \text{previous work} \\
&= \sum_{i=1}^{K^*} p_i(t) \frac{(1+\epsilon_i - \sum_{k=1}^{\infty} \epsilon_k p_k(t) + O(\|\epsilon\|_\infty^2))^j}{(j-1)!} \cdot \\
&\qquad \cdot \exp\left( - \left( 1 + \epsilon_i - \sum_{k=1}^{\infty} \epsilon_k p_k(t) + O(\|\epsilon\|_\infty^2) \right) \right); \\
&\qquad \text{Binomial expansion and Taylor expansion of exp} \\
&= \sum_{i=1}^{K^*} \frac{p_i(t)}{(j-1)!} \left( 1 + j \left( \epsilon_i - \sum_{k=1}^{\infty} \epsilon_k p_k(t) \right) + O(\|\epsilon\|_\infty^2) \right) \cdot \\
&\qquad \cdot e^{-1} \left( 1 - \left( \epsilon_i - \sum_{k=1}^{\infty} \epsilon_k p_k(t) \right) + O(\|\epsilon\|_\infty^2) \right) \\
&= \sum_{i=1}^{K^*} \frac{p_i(t)}{(j-1)!} \left( 1 + (j-1) \left( \epsilon_i - \sum_{k=1}^{\infty} \epsilon_k p_k(t) \right) + O(\|\epsilon\|_\infty^2) \right) e^{-1}; \\
&\qquad \text{as } \sum_{i=1}^{K^*} p_i = 1 \\
&= \frac{e^{-1}}{(j-1)!} + \frac{e^{-1}}{(j-2)!} \sum_{i=1}^{K^*} \left( \epsilon_i p_i(t) - p_i \sum_{k=1}^{\infty} \epsilon_k p_k(t) \right) + \frac{O(\|\epsilon\|_\infty^2)}{(j-1)!}; \\
&\qquad \text{series rearrangement} \\
&= \frac{e^{-1}}{(j-1)!} + \frac{e^{-1}}{(j-2)!} \left( \sum_{i=1}^{K^*} \epsilon_i p_i(t) - \sum_{k=1}^{\infty} \epsilon_k p_k(t) \sum_{i=1}^{K^*} p_i \right) + \frac{O(\|\epsilon\|_\infty^2)}{(j-1)!}; \\
&\qquad \text{as } \sum_{i=1}^{K^*} p_i = 1,\ p_{K^*}(t) = \sum_{k=K}^{\infty} p_k(t) \text{ and } \epsilon_k = \epsilon_K \text{ for all } k \geq K \\
&= \frac{e^{-1} + O(\|\epsilon\|_\infty^2)}{(j-1)!}.
\end{aligned}
$$

This completes the proof of the proposition. □

**Example 3.12.** We could emulate Sibert, Austerlitz and Heyer's model (§2.3) by using deterministic weights $w_j = j^\alpha$, for $1 \leq j < K$, and $w_j = K^\alpha$ for $j \geq K$.

When $\alpha$ and $K$ are small, the weights might be approximated by $1 + j\alpha$. So, by Proposition 3.11, the family size of a randomly picked individual is, in this case, $Pois(1) + 1$ up to an error $O((K\alpha)^2)$.

## 3.3 The impact of cultural transmission on family size

*Expectation, variance and trans-generational correlation*

We have seen so far that the distribution of the family size of a randomly chosen individual is well-approximated by a convex combination of Poisson distributions. In this section, with the help of the probability generating function, we find the moments of the family size. We then calculate the correlation between the family sizes of a randomly chosen individual and of its parent, and also give a way to recover the weight sequence $w$ by looking at both the family size and the parental family size distributions. These results may then be used to calibrate the model to observed data.

**Definition 3.13** (Family size)**.** *We write $S_t$ for the family size of a randomly picked individual from generation $t$ of a large but finite population evolving according to the CTF extension of the Wright-Fisher model with weight sequences of constant tail. With our notation for family size class proportion, we have $\mathbf{P}(S_t = j) = p_j(t)$. Let $g_{S_t} : \mathbf{R} \to \mathbf{R}$ denote the probability generating function of $S_t$.*

In the definition above both the random variable $S_t$ and the distribution $p(t)$ are functions of the total population size $N$, but in this section we suppress this dependence for ease of notation. One could think of $p(t)$ as $\lim_{N\to\infty} p^{(N)}(t)$, but we should avoid drawing from an infinite population.

As a first step, we use the recursion obtained in §3.1 for $p(t)$ to calculate the probability generating function $g_{S_{t+1}}$ conditional on the previous generation, $\mathcal{F}_t$.

**Proposition 3.14.** *The conditional probability generating function $g_{S_t}$ is defined on the whole real line, and we have $g_{S_{t+1}}(x) = \sum_{i=1}^{K*} p_i(t) r_i(t) x e^{r_i(t)(x-1)}$, for all $x \in \mathbf{R}$.*

*Proof.* We simply substitute Recursion (3.1) into the definition of the generating function of the family size.

$$\begin{aligned}
g_{S_{t+1}}(x) &= \mathrm{E}\left[x^{S_{t+1}}|\mathcal{F}_t\right] = \sum_{j=1}^{\infty} p_j(t+1)x^j; \text{ by (3.1)} \\
&= \sum_{j=1}^{\infty}\sum_{k=1}^{K^*} \frac{p_k(t)r_k(t)^j}{(j-1)!}x^j e^{-r_k(t)}; \text{ substituting } i \text{ for } j-1 \\
&= \sum_{k=1}^{K^*}\sum_{i=0}^{\infty} \frac{p_k(t)r_k(t)^{i+1}}{i!}x^{i+1}e^{-r_k(t)} \\
&= \sum_{k=1}^{K^*} p_k(t)r_k(t)xe^{-r_k(t)}\sum_{i=0}^{\infty}\frac{r_k(t)^i}{i!}x^i; \\
&\quad \text{by Taylor expansion of exp which converges for all } x \in \mathbf{R} \\
&= \sum_{k=1}^{K^*} p_k(t)r_k(t)xe^{r_k(t)(x-1)}.
\end{aligned}$$

□

**Remark 3.15.** Note that the probability generating function of a random variable $X \sim Pois(r)$ is $\exp(r(x-1))$, and hence of $X+1$ is $x\exp(r(x-1))$. We could have therefore deduced this result by referring to the fact that the distribution of the family size is a convex combination of $Pois + 1$ distributions.

Next, we use the probability generating function $g_{S_t}$ to recover the moments of the family size distribution.

**Proposition 3.16.** *The factorial moments of $S_{t+1}$, the family size of a randomly picked individual in generation $t+1$, exist; and for all natural numbers $t$ and $n$, we have*

$$\mathrm{E}\left[(S_{t+1})_{(n)}|\mathcal{F}_t\right] = \sum_{m=1}^{K^*} p_m(t)(r_m(t)^{n+1}(t) + nr_m(t)^n),$$

*where the binomial polynomial of degree n, $(x)_n$, is defined by $x_{(n)} = x(x-1)\cdots(x-(n-1))$.*

The $x_{(n)}$'s are also known as the falling factorials. The proposition is proved by differentiating the conditional probability generating function for $S_{t+1}$ $n$ times.

*Proof.* Since $g_{S_{t+1}}$ is a product of a polynomial and an exponential, it is infinitely many times differentiable everywhere. For all $n, t \in \mathbf{N}$, we obtain that

$$\begin{aligned} g^{(n)}_{S_{t+1}}(x) &= \sum_{m=1}^{K^*} p_m(t) r_m(t) \left(x e^{r_m(t)(x-1)}\right)^{(n)}; \text{ by Liebniz rule for derivatives} \\ &= \sum_{m=1}^{K^*} p_m(t) r_m(t) \left( \binom{n}{0} x r_m(t)^n + \binom{n}{1} r_m(t)^{n-1} \right) e^{r_m(t)(x-1)} \\ &= \sum_{m=1}^{K^*} p_m(t)(r_m(t)^{n+1} x + n r_m(t)^n) e^{r_m(t)(x-1)}. \end{aligned}$$

In particular, the $n$th conditional factorial moment given $\mathcal{F}_t$, $\mathrm{E}\,(S_{t+1})_{(n)} = g^{(n)}_{S_{t+1}}(1-)$ exists for all $n \in \mathbf{N}$ and equals $\sum_{m=1}^{K^*} p_m(t)(r_m(t)^{n+1}(t) + n r_m(t)^n)$. □

**Corollary 3.17.** *The expected family size of a randomly drawn individual is*

$$\mathrm{E}\,[S_{t+1}|\mathcal{F}_t] = 1 + \sum_{m=1}^{K^*} p_m(t) r_m(t)^2,$$

*while the variance of the family size is*

$$\mathrm{var}\,[S_{t+1}|\mathcal{F}_t] = \sum_{m=1}^{K^*} p_m(t)(r_m(t)^3 + r_m(t)^2) - \left( \sum_{m=1}^{K^*} p_m(t) r_m(t)^2 \right)^2.$$

*Proof.* Note that, in any generation $t$, $\sum_{m=1}^{K^*} p_m(t) r_m(t) = 1$ and that $\mathrm{var}\,(S_{t+1}) = \mathrm{E}\,S_{t+1}(S_{t+1}-1) + \mathrm{E}\,S_{t+1} - (\mathrm{E}\,S_{t+1})^2 = g''(1-) + g'(1-) - g'(1-)^2$, where expectation is conditional on the previous generation, $\mathcal{F}_t$. □

We now turn to the correlation between the family size of a randomly picked individual from the population and of its parent. So that we can speak about parental family size, we make the following definition.

**Definition 3.18.** *Write $S_t^*$ for the family size of the parent of a randomly selected individual in generation $t$.*

In order to calculate the correlation, we will need the distribution of the parents' family size, and also the joint distribution of $S_t$ and $S_t^*$.

**Proposition 3.19.** *The joint distribution of $S_{t+1}$ and $S_{t+1}^*$ (conditional on generation $t$), for all $i, j \in \mathbf{N}^+$, is given by*

$$\mathbf{P}\{S_{t+1} = j, S_{t+1}^* = i | \mathcal{F}_t\} = p_i(t) \frac{r_i(t)^j}{(j-1)!} e^{-r_i(t)},$$

*and hence the marginal distribution of $S_{t+1}^*$ is*

$$\mathbf{P}\{S_{t+1}^* = i | \mathcal{F}_t\} = p_i(t) r_i(t).$$

*Proof.* For the joint distribution of $S_{t+1}$ and $S_{t+1}^*$, note that the number of individuals of family size $j$ in generation $t+1$ whose parents have family size $i$ is $\sum_{l=1}^{Np_i^{(N)}(t)} j\, I_l^{(i)(j)}$, using the notation introduced in the proof of Theorem 3.2. Now taking conditional expectation with respect to $\mathcal{F}_t$, dividing by the population size, and finally taking limit as $N$ becomes large, we obtain the result.

The distribution of the parent's family size is obtained by summing over all

possible values of the randomly picked individual's family size:

$$\begin{aligned}
\mathbf{P}\{S^*_{t+1} = i|\mathcal{F}_t\} &= \sum_{j=1}^{\infty} \mathbf{P}\{S_{t+1} = j, S^*_{t+1} = i|\mathcal{F}_t\}; \text{by first part} \\
&= \sum_{j=1}^{\infty} p_i(t) \frac{r_i(t)^j}{(j-1)!} e^{-r_i(t)}; \text{ change of variable } j \to j+1 \\
&= p_i(t) r_i(t) \sum_{j=0}^{\infty} \frac{r_i(t)^j}{j!} e^{-r_i(t)}; \text{ Taylor expansion of exp} \\
&= p_i(t) r_i(t).
\end{aligned}$$

□

**Remark 3.20.** This result can be exploited to recover the weights $w$ behind the model just by looking at the family size and the parental family size distributions. We obtain the relative infinitesimal weights $r$ by $r_i(t) = \mathbf{P}\{S^*_{t+1} = i|\mathcal{F}_t\}/\mathbf{P}\{S_t = i\}$. Now the underlying weights are any of the non-zero solutions to the system of linear equations:

$$w_i = r_i(t) \sum_{j=1}^{K^*} p_j(t) w_j,$$

for all $1 \leq j \leq K$. (There are $K - 1$ independent and consistent equations, so the solution to the homogeneous system is a one-dimensional subspace of $\mathbf{R}^K$.)

**Proposition 3.21.** *The moments of the parent's family size are expressible in terms of the moments of the family size of the child and first $K$ elements of the sequences $(p_j)$ and $(r_j)$. We have, for all $t \in \mathbf{N}^+$,*

$$\mathrm{E}\, S^*_{t+1} = r_K(t) \mathrm{E}\, S_t + \sum_{j=1}^{K-1} j(r_j(t) - r_K(t)) p_j(t),$$

$$\mathrm{E}\, {S^*_{t+1}}^2 = r_K(t)\mathrm{E}\, S_t^2 + \sum_{j=1}^{K-1} j^2(r_j(t) - r_K(t))p_j(t)$$

*and that the mixed second moment,* $\mathrm{E}\, S_t S_t^*$*, is the sum of the first moments. Expectations are conditional on generation* $t$*,* $\mathcal{F}_t$*.*

*Proof.* As for the moments of $S^*_{t+1}$, note that the tail of the sequence $(r_j(t))$, the infinitesimal relative weights, is constant: for all $t \in \mathbf{N}^+$ and $j \geq K$ we have $r_j(t) = r_K(t)$. To obtain the mixed second moment, we rearrange the absolutely convergent series:

$$\begin{aligned}
\mathrm{E}\, S_{t+1}S^*_{t+1} &= \sum_{i,j=1}^{\infty} ij\mathbf{P}\{S_{t+1} = j, S^*_{t+1} = i\};\ \text{by Proposition 3.19} \\
&= \sum_{i,j=1}^{\infty} ijp_i(t)\frac{r_i(t)^j}{(j-1)!}e^{-r_i(t)};\ \text{rearrangement} \\
&= \sum_{i=1}^{\infty} ir_i(t)p_i(t)\sum_{j=1}^{\infty} j\frac{r_i(t)^{j-1}}{(j-1)!}e^{-r_i(t)};\ \text{Taylor expansion} \\
&= \sum_{i=1}^{\infty} ir_i(t)(1 + r_i(t))p_i(t) \\
&= \sum_{i=1}^{K-1} i(r_i(t)^2 - r_K(t)^2)p_i(t) + r_K(t)^2\mathrm{E}\, S_t + \mathrm{E}\, S^*_{t+1}.
\end{aligned}$$

Note that the series in line 2 is convergent because of the Cauchy-Schwarz inequality, and it is absolutely convergent as $r$ is a bounded sequence. □

As we have seen in §1.2 the trans-generational correlation in family size is negative in some human populations. Below we give a condition for the level of correlation to be non-negative.

**Corollary 3.22.** *The sign of the correlation between* $S$ *and* $S^*$ *is determined by*

*their first moments. Moreover,* $\operatorname{cov}(S_{t+1}, S^*_{t+1}) \geq 0$ *if, and only if*

$$\frac{\sum_{i=1}^{K-1} i(r_i(t)^2 - r_K(t)^2)p_i(t) + r_K(t)^2 \mathrm{E}\, S_t}{\mathrm{E}\, S_{t+1} - 1} \geq \mathrm{E}\, S^*_{t+1}.$$

We note that $\mathrm{E}\, S^*$ and $\mathrm{E}\, S$ have already been expressed in terms of $(p_i)$ in Corollary 3.17 and Proposition 3.21.

*Proof.* Using Proposition 3.21, for all generation $t$, we have

$$\begin{aligned}
\operatorname{cov}(S_{t+1}, S^*_{t+1}) &= \mathrm{E}\, S_{t+1}S^*_{t+1} - \mathrm{E}\, S_{t+1}\mathrm{E}\, S^*_{t+1} \\
&= \sum_{i=1}^{K-1} i(r_i(t)^2 - r_K(t)^2)p_i(t) + r_K(t)^2 \mathrm{E}\, S_t + \mathrm{E}\, S^*_{t+1} - \mathrm{E}\, S_{t+1}\mathrm{E}\, S^*_{t+1} \\
&= \sum_{i=1}^{K-1} i(r_i(t)^2 - r_K(t)^2)p_i(t) + r_K(t)^2 \mathrm{E}\, S_t + (1 - \mathrm{E}\, S_{t+1})\mathrm{E}\, S^*_{t+1} \\
&\geq 0 \quad \text{if and only if}
\end{aligned}$$

$$\mathrm{E}\, S^*_{t+1}(\mathrm{E}\, S_{t+1} - 1) \leq \sum_{i=1}^{K-1} i(r_i(t)^2 - r_K(t)^2)p_i(t) + r_K(t)^2 \mathrm{E}\, S_t$$

Dividing through by $\mathrm{E}\, S_{t+1} - 1$ does not change the direction of the relation, as a family consists of at least one member, and hence $\mathrm{E}\, S_{t+1} \geq 1$. □

**Remark 3.23.** The trans-generational correlation itself is a rather involved expression, $\frac{\operatorname{cov}(S_{t+1}, S^*_{t+1})}{\sqrt{\operatorname{var} S_{t+1} \operatorname{var} S^*_{t+1}}}$, but an explicit formula may be obtained in terms of the first $K$ terms of the infinitesimal relative weight sequence $r$ combining Corollary 3.17 and Proposition 3.21.

In Example 3.10 we gave a specific weight sequence which causes the time stationary family size distribution to flatten around family size 5, reminiscent of the observed distribution in the Saguenay-Lac-Saint-Jean population. We illustrate the

section's results by calculating expectations, variances of and correlation between the family size and the parental family size for these weights.

**Example 3.24.** Setting the weight sequence $w = (1/2, 1, 16, 1, 4, 4, 1, 1, ...)$, as in Example 3.10, the expected time stationary family size $\mathrm{E}\, S \approx 4.22$ and the variance of the family size $\mathrm{var}\, S \approx 7.78$. The expected parental family size $\mathrm{E}\, S^* \approx 3.94$, while its variance $\mathrm{var}\, S^* \approx 3.19$. Finally, the trans-generational correlation between family sizes $S$ and $S^*$, is 0.3374. Compare this value to the observed value of 0.21 today and [sic] 0.34 a century ago in the Saguenay-Lac-Saint-Jean region. The two marginal distributions are plotted together in Figure 4.

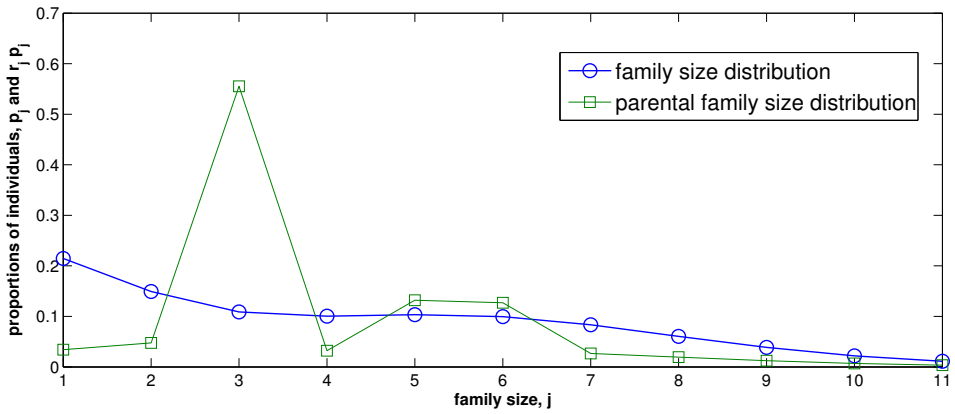



Figure 4: The distribution of a randomly picked individual's family size (circles), and the distribution of the family size of a randomly picked individual's parent (squares), for the weight sequence $w = (1/2, 1, 16, 1, 4, 4, 1, 1, ...)$.

Another example is the perturbation of $w$ about the Wright-Fisher case.

**Example 3.25.** Assuming that the weight sequence $w = (1 + \epsilon_j)$, with $\epsilon$ being of constant tail and $\|\epsilon\|_\infty > 0$ small, the family size distribution $p$ is time-invariant.

The expected family size is

$$\begin{aligned}
\mathrm{E}[S] &= \sum_{j=1}^{\infty} j p_j = \sum_{j=1}^{\infty} j \frac{e^{-1}}{(j-1)!} + O(\|\epsilon\|^2) \\
&= \sum_{j=0}^{\infty} (j+1) \frac{e^{-1}}{j!} + O(\|\epsilon\|^2) = 1 + 1 + O(\|\epsilon\|^2) \\
&= 2 + O(\|\epsilon\|^2).
\end{aligned}$$

While the expected parental family size is

$$\begin{aligned}
\mathrm{E}[S^*] &= \sum_{j=1}^{\infty} j p_j r_j \\
&= \sum_{j=1}^{\infty} (1+\epsilon_j) \frac{j}{(j-1)!} e^{-1} - \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \frac{\epsilon_i}{(i-1)!} e^{-1} \frac{j}{(j-1)!} e^{-1} + O(\|\epsilon\|^2) \\
&= 2 + \sum_{j=1}^{\infty} \frac{j\epsilon_j}{(j-1)!} e^{-1} - \sum_{j=1}^{\infty} \frac{j}{(j-1)!} \sum_{i=1}^{\infty} \frac{\epsilon_i}{(i-1)!} e^{-1} + O(\|\epsilon\|^2) \\
&= 2 + \sum_{j=1}^{\infty} \frac{j\epsilon_j}{(j-1)!} e^{-1} - 2 \sum_{i=1}^{\infty} \frac{\epsilon_i}{(i-1)!} e^{-1} + O(\|\epsilon\|^2) \\
&= 2 + \sum_{j=1}^{\infty} (j-2) \frac{\epsilon_j}{(j-1)!} e^{-1} + O(\|\epsilon\|^2).
\end{aligned}$$

We calculate the covariance of $S$ and $S^*$ by finding their mixed second moment first.

$$
\begin{aligned}
\mathrm{E}\,[SS^*] &= \sum_{i,j=1}^{\infty} ij\mathbf{P}\{S^*=i, S=j\}\text{; by Proposition 3.19}\\
&= \sum_{i,j=1}^{\infty} ijp_i\frac{r_i^j}{(j-1)!}e^{-r_i}\text{; by Proposition \textbf{??}}\\
&= \sum_{i,j=1}^{\infty} \frac{ij}{(i-1)!(j-1)!}e^{-1}\frac{\left(1+\epsilon_i-\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}\right)^j}{(j-1)!}\cdot\\
&\qquad \cdot\exp\left(-1-\epsilon_i+\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}\right)+O(\|\epsilon\|^2)\text{; by Taylor's Theorem}\\
&= \sum_{i,j=1}^{\infty} \frac{ij}{(i-1)!(j-1)!}e^{-1}\left(1+j\epsilon_i-j\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}\right)e^{-1}\cdot\\
&\qquad \cdot\left(1-\epsilon_i+\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}\right)+O(\|\epsilon\|^2)\\
&= \sum_{i=1}^{\infty}\frac{i}{(i-1)!}e^{-1}\sum_{j=1}^{\infty}\frac{j}{(j-1)!}e^{-1}+\sum_{i=1}^{\infty}\frac{i\epsilon_i}{(i-1)!}e^{-1}\sum_{j=1}^{\infty}\frac{j(j-1)}{(j-1)!}^{-1}-\\
&\qquad -\sum_{i=1}^{\infty}\frac{i}{(i-1)!}e^{-1}\sum_{j=1}^{\infty}\frac{j^2}{(j-1)!}e^{-1}\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}\text{; change of variables}\\
&= 2\cdot 2+\sum_{j=0}^{\infty}\frac{j(j+1)}{j!}e^{-1}\sum_{k=1}^{\infty}\frac{k\epsilon_k}{(k-1)!}e^{-1}-\\
&\qquad -2\sum_{j=0}^{\infty}\frac{(j+1)^2}{j!}e^{-1}\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}+O(\|\epsilon\|^2)\\
&= 4+\sum_{j=1}^{\inf}\frac{j+1}{(j-1)!}e^{-1}\sum_{k=1}^{\infty}\frac{k\epsilon_k}{(k-1)!}e^{-1}-\\
&\qquad -2\sum_{j=0}^{\infty}\left(\frac{j^2}{j!}+2\frac{j}{j!}+\frac{1}{j!}\right)e^{-1}\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}+O(\|\epsilon\|^2)\\
&= 4+(1+2)\sum_{k=1}^{\infty}\frac{k\epsilon_k}{(k-1)!}e^{-1}-2((1+1)+2+1)\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}+O(\|\epsilon\|^2)\\
&= 4+3\sum_{k=1}^{\infty}\frac{k\epsilon_k}{(k-1)!}e^{-1}-10\sum_{k=1}^{\infty}\frac{\epsilon_k}{(k-1)!}e^{-1}+O(\|\epsilon\|^2)\\
&= 4+\sum_{k=1}^{\infty}(3k-10)\frac{\epsilon_k}{(k-1)!}e^{-1}+O(\|\epsilon\|^2).
\end{aligned}
$$

Hence their covariance is

$$\begin{aligned}
\operatorname{cov}[S, S^*] &= \mathrm{E}[S, S^*] - \mathrm{E}[S]\mathrm{E}[S^*];\ \text{using previous work} \\
&= \sum_{k=1}^{\infty}(3k-10)\frac{\epsilon_k}{(k-1)!}e^{-1} - 2\sum_{k=1}^{\infty}(k-2)\frac{\epsilon_k}{(k-1)!}e^{-1} + O(\|\epsilon\|^2) \\
&= \sum_{k=1}^{\infty}(k-6)\frac{\epsilon_k}{(k-1)!}e^{-1} + O(\|\epsilon\|^2)
\end{aligned}$$

This means that, for example, if the first five terms of the sequence $\epsilon$ are positive and the rest are negative, then the correlation between the family size of a randomly chosen individual and that of her parent is negative. In this situation, having at most five children is advantageous to varying degree, while a family size above five may place too much strain onto the mother.

## 3.4 The coalescent dual of the model

In this final section of the chapter, we show that, using an appropriate time scale, the ancestral process in our model converge to the $n$-coalescent (Definition 2.4) as the population size becomes large.

The $n$-coalescent approximates the genealogies of samples drawn from the population for many forward-in-time models used in population genetics. In §2.2, we showed that this is indeed the case for the Wright-Fisher model. However, the Moran, and some of the Cannings models have the $n$-coalescent as their time-rescaled dual limits, too. Moreover, provided that the sample size $n \ll N$, the $n$-coalescent approximates the *exact coalescent*, the pre-limiting coalescent dual of W-F really well. The expected time to the most recent common ancestor of the sample is almost the same in the two dual processes, (see Fu (2006) for details). In view of these remarks, it is not surprising that the $n$-coalescent is preferred in applications for estimating the behaviour of the genealogies of samples drawn from the population. It is therefore important to know whether the ancestral process in our extension tend to the same limit as that of W-F. We shall establish this in the case of time-stationary family size distributions by closely following the collapse of structure argument in Nagylaki (1980), and Nordborg and Krone (2002). Ancestral lines in our model jump between family size classes at each generation, while they rarely coalesce (genealogies of pairs merge with probability $O(1/N)$, multiple lines with probability $O(1/N^2)$). So, we expect to see a 'separation of time-scales', that is, measuring time in units of $N$ the information of the family size structure vanishes in the sense that the detail is replaced by a single statistic, the number of surviving lineages.

Our aim is to trace the ancestry of a sample taken from the population. The

technique is based on, instead of the basic ancestral process, a more refined process, the backwards evolution of configurations. This configuration process encodes the number of the ancestors classified by their family sizes.

**Definition 3.26** (The backward configuration process)**.** *Take a sample of $m$ individuals from the time stationary distribution of the forward model. Let the random variable $X_j(t)$ denote the number of ancestors, in generation $t$, of the sample, whose family size is $j$. The configuration process $X(t)$ of the ancestors is then defined by $X(t) = (X_1(t), X_2(t), ...X_{K^*}(t))$. Here, the index $K^*$ has the same meaning as before. $X_{K^*}(t)$ denotes the number of all ancestors in generation $t$ whose family size is* at least *$K$. The state space of X(t), the set of all possible configurations, is therefore $R = \bigcup_{n=1}^{m} R_n$, where $R_n$ is the set of all possible configurations with $n$ surviving lineages, that is, $R_1 = \{(1,0,0,...0), (0,1,0,...0),\ ...(0,0,0,...1)\}$; $R_2 = \{(2,0,0,...0), (0,2,0,...0),\ ...\ (0,0,0,...2), (1,1,0,...0)\ (1,0,1,...0), ...\}$ and so on. Time passes backwards.*

As suggested above, we will be interested in what happens in between coalescent events. To this end, we define the $n$-level configuration process, the process when there are $n$ individuals ancestral to the sample surviving at generation $t$.

**Definition 3.27** (The $n$-level configuration process)**.** *Let the random variable $X^{(n)}(t)$ be the restriction of the backward configuration process $X(t) = (X_1(t), X_2(t), ...X_K^*(t))$ to the time interval when $\sum_{j=1}^{K^*} X_j(t) = n$. Whenever a coalescent event happens, the process $X^{(n)}(t)$ stops.*

The processes $X$ and $X^{(n)}$ are Markov chains; and we adopt the terminology usually associated with them. We say that a line is in state $i$ if the ancestor in the current generation on the line has family size $i$. It migrates from state $i$ to state $j$, if the parent of an individual of family size $i$ has family size $j$. We will deem two lines

to be coalesced whenever the lines are in the same family size class and they choose the same family. Although the lines would coalesce only in the next generation, this convention makes the calculation of the probability that two lines coalesce in the previous generation easier.

We denote the pre-limiting backward migration process of a single lineage by $S_N(t)$. Thus $S_N(t)$ is the family size of the ancestor living in generation $t$. The state space of this Markov chain is $\{1, 2, 3, \ldots N\}$; and below we calculate its transition matrix. Once we have the transition rates for $S_N$, we will use them in writing down the transition matrix of the whole configuration process $X_N$.

**Proposition 3.28.** *The backward migration process is a Markov chain with transition probabilities $b_{ij}^{(N)}$ from state $i$ to state $j$ given by*

$$b_{ij}^{(N)} = \frac{p_j}{p_i} \frac{{r_j}^i}{(i-1)!} e^{-r_j} + \\ + \left( \frac{\bar{p}_j}{p_i} r_j^i - \frac{p_j}{p_i^2} \bar{p}_i r_j^i + i \frac{p_j}{p_i} \bar{r}_j r_j^{i-1} - \frac{p_j}{p_i} \bar{r}_j r_j^i \right) \frac{e^{-r_j}}{N} + O\left( \frac{1}{N^2} \right).$$

*As a consequence, the process is irreducible and aperiodic and hence has an invariant measure $\gamma^{(N)}$ which satisfies the relation*

$$\gamma_i^{(N)} = \sum_{j=1}^{N} \gamma_j^{(N)} \frac{p_i}{p_j} \frac{{r_i}^j}{(j-1)!} e^{-r_i} + O\left( \frac{1}{N} \right) \tag{3.7}$$

*and $\gamma^{(N)} = \gamma + O(1/N)$, where $\gamma$ is the unique invariant measure of the Markov chain with one step transition probabilities*

$$b_{ij} = \frac{p_j}{p_i} \frac{{r_j}^i}{(i-1)!} e^{-r_j}.$$

The quantities $p_j$ and $\bar{p}_j$ are the infinite-time limits of the sequences given by

recursion in Theorem 3.2 and Proposition **??**, respectively. In turn, $b_{ij}^{(N)}$ is the probability that an individual of family size $i$ has a parent whose family size is $j$.

*Proof.* In the proof of Theorem 3.2, we found the expected number of individuals of family size $i$ in generation $t+1$ whose parents are of family size $j$, conditional on $p^{(N)}(t)$, to be

$$Np_j^{(N)}(t)i\binom{N}{i}\left(\frac{r_j^{(N)}(t)}{N}\right)^i\left(1-\frac{r_j^{(N)}(t)}{N}\right)^{N-i}.$$

Writing $M_{ij}$ for the expected number of these individuals and noting that, within each family size class, the individuals are equally likely to reproduce and that $p^{(N)}(t)$

is assumed to be time-invariant, we obtain that

$$
\begin{aligned}
b_{ij}^{(N)}(t) &= \frac{M_{ij}}{\sum_{k=1}^{\infty} M_{ik}} = \frac{N p_j^{(N)}(t) i \binom{N}{i} \left(\frac{r_j^{(N)}(t)}{N}\right)^i \left(1 - \frac{r_j^{(N)}(t)}{N}\right)^{N-i}}{N p_i^{(N)}(t+1)} \\
&= \frac{p_j(t) + \frac{\bar{p(t)}_j}{N} + O(N^{-2})}{p_i(t+1) + \frac{\bar{p}_i(t+1)}{N} + O(N^{-2})} \left(r_j(t) + \frac{\bar{r}_j(t)}{N} + O(N^{-2})\right)^i e^{-r_j(t) - \frac{\bar{r}_j(t)}{N} + O(N^{-2})}; \\
&\qquad \text{as the sample was taken from the time stationary distribution,} \\
&\qquad \text{we have } p_i(t+1) = p_i(t) = p_i \\
&= \left(p_j + \frac{\bar{p}_j}{N}\right) \frac{1}{p_i} \frac{1}{1 + \left(\frac{\bar{p}_i}{p_i} + O(N^{-1})\right)\frac{1}{N}} \left(r_j^i + i\frac{\bar{r}_j r_j^{i-1}}{N}\right) e^{-r_j} \left(1 - \frac{\bar{r}_j}{N}\right) + \\
&\qquad + O(N^{-2}); \\
&\qquad \text{taking the geometric expansion of the third factor } (N \text{ is large}) \\
&\qquad \text{and grouping powers of } N \\
&= \frac{p_j}{p_i} \frac{r_j^i}{(i-1)!} e^{-r_j} + \\
&\qquad + \left(\frac{\bar{p}_j}{p_i} r_j^i - \frac{p_j}{p_i^2} \bar{p}_i r_j^i + i \frac{p_j}{p_i} \bar{r}_j r_j^{i-1} - \frac{p_j}{p_i} \bar{r}_j r_j^i\right) \frac{e^{-r_j}}{N} + O\left(\frac{1}{N^2}\right) \\
&= b_{ij} + \frac{\bar{b}_{ij}}{N} + O\left(\frac{1}{N^2}\right),
\end{aligned}
\tag{3.8}
$$

defining $b_{ij} = \frac{p_j}{p_i} \frac{r_j^i}{(i-1)!} e^{-r_j}$, and the coefficient $\bar{b}_{ij}$ correspondingly.

Using the $*$-convention, we define the $K$ by $K$ one-step transition matrix $B_n^{(N)}$ of the $n$th level backward migration process by $B_n^{(N)} = (b_{ij}^{(N)})$, where $b_{iK^*}^{(N)} = 1 - \sum_{j=1}^{K-1} b_{ij}^{(N)}$ and $b_{K^*j}^{(N)} = 1 - \sum_{i=1}^{K-1} b_{ij}^{(N)}$, assuming that $N \geq K$. We may also define $B_n = (b_{ij})$, with entries in the last column $b_{iK^*} = 1 - \sum_{j=1}^{K-1} b_{ij}$ and entries in the last row $b_{K^*j} = 1 - \sum_{i=1}^{K-1} b_{ij}$. The matrix $\bar{B}_n$ is defined likewise. Hence the three

matrices are related by the equation

$$B_n^{(N)} = B_n + N^{-1}\bar{B}_n + O(N^{-2}).$$

Next we observe that the backward transition probabilities $b_{ij}^{(N)} > 0$, for all family size class $1 \leq i, j \leq K^*$. As a result, any state can be reached from any other one in exactly one step. Hence the Markov chain is irreducible and aperiodic: it has a unique invariant measure $\gamma^{(N)}$. Analogously, $b_{ij} > 0$ for all $i, j \in \mathbf{N}^+$, and hence the process given by these transition probabilities is ergodic, too. Let $\gamma$ denote its unique invariant measure. We should like to approximate $\gamma^{(N)}$ by $\gamma$. As $\gamma^{(N)}$ is invariant, for all $i \in \mathbf{N}^+$, we have

$$\begin{aligned}
\gamma_i^{(N)} &= \sum_{j=1}^{\infty} \gamma_j^{(N)} b_{ij}^{(N)};\ \text{substituting (3.8)} \\
&= \sum_{j=1}^{\infty} \gamma_j^{(N)} (b_{ij} + O(N^{-1})) \\
&= \sum_{j=1}^{\infty} \gamma_j^{(N)} + O(N^{-1}) + \sum_{j=1}^{\infty} \gamma_j^{(N)} b_{ij};\ \text{as} \sum_{j=1}^{\infty} \gamma_j^{(N)} = 1 \\
&= \sum_{j=1}^{\infty} \gamma_j^{(N)} b_{ij} + O(N^{-1});
\end{aligned}$$

Only the assertion $\gamma^{(N)} = \gamma + O(1/N)$ remains to be proved. Here we show this only for the finite dimensional version of $\gamma$ and $\gamma^{(N)}$: we truncate the sequences $\gamma^{(N)}$ and $\gamma$ by defining their $K$th term to be $\gamma_{K^*} = 1 - \sum_{j=1}^{K-1} \gamma_j$ and $\gamma_{K^*}^{(N)} = 1 - \sum_{j=1}^{K-1} \gamma_j^{(N)}$, respectively. Let $I_K$ denote the $K \times K$ identity matrix. Because $\gamma$ is an invariant measure for B, we have $\gamma^{(N)}(I - B^{(N)}) = 0$ and $\gamma(I - B) = 0$. Thus we obtain that

$$(\gamma - \gamma^{(N)})(I - B) = O(N^{-1}),$$

where $O(1/N)$ denotes a vector whose entries are of order $1/N$. As $\gamma$ in the unique invariant measure corresponding to the transition matrix $B$, $\gamma - \gamma^{(N)} = O(1/N)$.

We shall complete the proof for indices greater than $K$ in Remark 3.31. □

As we have seen the invariant measure $\gamma$ is the unique solution of the integral equation (3.7) up to an error of order $O(1/N)$. It is easy to find $\gamma$ in the Wright-Fisher case, that is, when the weight sequence $w$ is a positive constant multiple of $(1, 1, 1, \dots)$.

**Corollary 3.29.** *The limit invariant measure $\gamma$ is given by $(\gamma_j) = (p_j) = \left(\frac{e^{-1}}{(j-1)!}\right)$ if and only if the infinitesimal relative weights $(r_j) = (1, 1, 1, ...)$ or, equivalently, the reproductive weights are $(w_j) = (a, a, a, ...)$ with $a \in \mathbf{R}^+$.*

*Proof.* To see this, we note that, substituting $p$ for $\gamma$ into the integral equation, we have, for all $j \in \mathbf{N}^+$,

$$\begin{aligned} p_j &= \sum_{i=1}^{\infty} p_j \frac{p_i}{p_j} \frac{r_j^i}{(i-1)!} e^{-r_j} = \sum_{i=1}^{\infty} p_j \frac{r_j^i}{(i-1)!} e^{-r_j} \\ &= r_j p_j \sum_{i=1}^{\infty} \frac{r_j^{i-1}}{(i-1)!} e^{-r_j}; \text{ change of variable and Taylor expansion} \\ &= r_j p_j. \end{aligned}$$

The terms of the sequence $p$ are non-zero, thus $p$ is a solution of the integral equation if and only if $(r_j) = (1, 1, 1, ...)$. Substituting $r$ back, we get that $p_j = \frac{e^{-1}}{(j-1)!}$. Finally, knowing that $(r_j) = (1, 1, 1, ...)$ we deduce from the definition of the relative infinitesimal weights:

$$r_j = \frac{w_j}{\sum_{i=1}^{\infty} w_i p_i}$$

that $(w_j) = a\,(1, 1, 1, ...)$ for all $a > 0$. □

This result will be useful in testing the validity of our formula for the appropriate time scale in which we lose information about the family size structure and have weak convergence to the $n$-coalescent. Furthermore, were we interested in only small deviations from the case of independent family sizes, we could perturb Equation (3.7) around the known solution in the W-F case, as we did in Proposition 3.11, to obtain an estimate for the invariant measure $\gamma$.

**Proposition 3.30.** *If the weights are of the form $w_i = 1 + \epsilon_i$, then the invariant measure $\gamma(\epsilon)$ of the backward migration process is given by $\gamma_i(\epsilon) = (1 + \epsilon_i)p_i(\epsilon) + O(\|\epsilon\|_\infty^2)$ for all $i \in \mathbf{N}^+$.*

*Proof.* The substitution $x = \frac{\gamma}{p}$ makes equation (3.7) more transparent:

$$x_i = \sum_{j=1}^{\infty} x_j \frac{r_i^j}{(j-1)!} e^{-r_i}.$$

Second, we recall from Proposition 3.11 that if $w_i = 1 + \epsilon_i$, for all $i$, $1 \leq i \leq K$, then we have

$$r_i(\epsilon) = 1 + \epsilon_i - \sum_{j=1}^{\infty} \alpha_j,$$

where we have used the substitution $\alpha_j = \epsilon_j p_j = \frac{\epsilon_j e^{-1} + O(\|\epsilon\|_\infty^2)}{(j-1)!}$ in order to simplify the notation. Since $x_i = 1$ in the Wright-Fisher case, we make the ansatz that $x_i(\epsilon) = 1 + f_i(\epsilon)$, for some function $f : \mathbf{R}^K \to \mathbf{R}$ to be determined. Rewriting the

integral equation with these substitutions, we obtain

$$\begin{aligned} 1 + f_i(\epsilon) &= \sum_{j=1}^{\infty} (1 + f_j(\epsilon)) \frac{r_i(\epsilon)^j}{(j-1)!} e^{-r_i(\epsilon)} \\ &= r_i(\epsilon) + \sum_{j=1}^{\infty} f_j(\epsilon) \frac{r_i(\epsilon)^j}{(j-1)!} e^{-r_i(\epsilon)}. \end{aligned}$$

Subtracting 1 from both sides of the equation yields

$$\begin{aligned} f_i(\epsilon) &= \epsilon_i - \sum_{k=1}^{\infty} \alpha_k + \sum_{j=1}^{\infty} f_j(\epsilon) \frac{(1 + \epsilon_i - \sum_{k=1}^{\infty} \alpha_k)^j}{(j-1)!} e^{-1-\epsilon_i+\sum_{k=1}^{\infty} \alpha_k} \\ &= \epsilon_i - \sum_{k=1}^{\infty} \alpha_k + \sum_{j=1}^{\infty} f_j(\epsilon) \frac{1 + j\left(\epsilon_i - \sum_{k=1}^{\infty} \alpha_k + O\left(\|\epsilon\|_\infty^2\right)\right)}{(j-1)!} \cdot \\ &\qquad \cdot \left(1 - \epsilon_i + \sum_{k=1}^{\infty} \alpha_k + O\left(\|\epsilon\|_\infty^2\right)\right) e^{-1} \\ &= \epsilon_i - \sum_{k=1}^{\infty} \alpha_k + \sum_{j=1}^{\infty} f_j(\epsilon) \left( \frac{1 + (j-1)(\epsilon_i - \sum_{k=1}^{\infty} \alpha_k)}{(j-1)!} + \right. \\ &\qquad \left. + \frac{O((f_j(\epsilon)\|\epsilon\|_\infty)}{(j-1)!} + \frac{O\left(\|\epsilon\|_\infty^2\right)}{(j-1)!} \right) e^{-1} \\ &= \epsilon_i - \sum_{k=1}^{\infty} \alpha_k + \sum_{j=1}^{\infty} \frac{f_j(\epsilon)}{(j-1)!} e^{-1} + O((\|f(\epsilon)\|_\infty \|\epsilon\|_\infty) + O\left(\|\epsilon\|_\infty^2\right), \end{aligned} \tag{3.9}$$

for all $i \in \mathbf{N}^+$. Substituting expression $\frac{\epsilon_k e^{-1} + O(\|\epsilon\|_\infty^2)}{(k-1)!}$ for $\alpha_k$ in equation (3.9), we see that $f_i(\epsilon) = \epsilon_i$ is a solution. This completes the proof, because $\gamma(\epsilon)$ and $p(\epsilon)$ are unique sequences and hence so is their ratio, $x(\epsilon)$. □

Below we briefly describe how these results may be applied to estimate the reproductive weights $w = (1 + \epsilon_i)$, where the $\epsilon_i$'s are small. First, one would observe the time stationary distribution of the family sizes, $p$, by looking at the present population. Using $p$ one calculates the infinitesimal relative weights $r$ by substituting

$p$ and $w$ into their definition. Knowing $r$, we solve the linear system of $K$ equations

$$\epsilon_i - \sum_{j=1}^{K} \frac{e^{-1}}{(j-1)!}\epsilon_j = r_i - 1,$$

for $\epsilon$; and hence we obtain an estimate for the underlying weights $w_i = 1 + \epsilon_i$. Proposition 3.30 now yields an estimate for the invariant measure $\gamma$ of the backward migration process.

**Remark 3.31.** How does the fact that the weight sequence $w$ has a constant tail affect the distribution $\gamma$? As $r_i = r_K$ for all $i \geq K$, adopting the notation $x_i = \gamma_i/p_i$ from the proof of Proposition 3.30, we also have

$$x_i = \sum_{j=1}^{\infty} x_j \frac{r_i^j}{(j-1)!} e^{-r_i} = \sum_{j=1}^{\infty} x_j \frac{r_K^j}{(j-1)!} e^{-r_K} = x_K,$$

for all $i \geq K$. So $\gamma_i = x_K\, p_i$, for all $i \geq K$; $\gamma$ and $p$ have proportional tails. In particular, this means that

$$\gamma_i^{(N)} = x_K^{(N)} p_i^{(N)} = (x_K + O(N^{-1}))(p_i + O(N^{-1})) = x_K p_i + O(N^{-1}) = \gamma_i + O(N^{-1}),$$

for all $i \geq K$, completing the proof of Proposition 3.28. It also follows that the sequence $x$ is bounded above, which we shall use later.

Exploiting that the sequence $x$, too, has a constant tail we may find its terms

by solving the following system of linear equations

$$\begin{aligned} x_i &= \sum_{j=1}^{\infty} x_j \frac{r_i^j}{(j-1)!} e^{-r_i} \\ &= \sum_{j=1}^{K} x_j \frac{r_i^j}{(j-1)!} e^{-r_i} + r_K x_K \sum_{j=K+1}^{\infty} \frac{r_K^{j-1}}{(j-1)!} e^{-r_K}, \end{aligned}$$

for all $1 \leq i \leq K$. Observe that the second term does not depend on $i$. Subtracting the equations from each other pairwise, therefore yields a linear system of equations:

$$x_{i+1} - x_i = \sum_{j=1}^{K} x_j \frac{r_{i+1}^j e^{-r_{i+1}} - r_i^j e^{-r_i}}{(j-1)!},$$

for all $i$, $1 \leq i \leq K-1$. If the first $K$ weights are distinct then this is a homogeneous system of $K-1$ independent and consistent linear equations with K unknowns. This means that the solution space is one-dimensional, that is, if we find one solution, we can be certain that after re-normalisation of $p_i x_i$ we get the correct terms for the invariant measure $\gamma$.

Once we have obtained the invariant measure for the backward migration process of the lineages, writing down an approximate invariant measure of the whole $n$th level configuration process is easy. We should be wary here of two things. First, the $n$th level process terminates at a coalescent event and such events happen with probability $O(1/N)$. Nevertheless, in our calculations, as we shall see later, we need the $O(1)$ term only of the invariant measure. Second, we need to establish the 'quasi' independence of the ancestral lines.

**Proposition 3.32.** *The nth level configuration process $X_N^{(n)}(t)$ has an invariant*

*measure* $\pi_n^{(N)}$*, given by*

$$\pi_n^{(N)}(x) = \frac{n!}{x_1! \dots x_{K^*}!} \gamma_1^{x_1} \dots \gamma_{K^*}^{x_{K^*}} + O(N^{-1})$$

*for all* $n$*th level configuration vector* $x \in R_n$*.*

We could calculate the $O(1/N)$ term of $\pi_n^{(N)}$, just as we did for the backward migration probabilities $b_{ij}^{(N)}$, but we will need the $O(1)$ term of $\pi_n^{(N)}$ only.

*Proof.* First we show that a pair of lines migrate almost independently of each other. Recall from the proof of Theorem 3.2 that the probability that two lineages migrate from state $i$ to $j$ and state $m$ to $k$, respectively, is

$$\begin{aligned}
&b_{ij}^{(N)} \frac{p_{kj}^{(N)}(t)}{p_{mi}^{(N)}(t+1)} m \binom{N}{m} \left( \frac{r_{kj}^{(N)}(t)}{N} \right)^m \left( 1 - \frac{r_{kj}^{(N)}(t)}{N} \right)^{N-m} \\
&= b_{ij} \frac{p_k(t)}{p_m(t+1)} \frac{r_{kj}(t)^m}{(m-1)!} e^{-r_{kj}(t)} + O(N^{-1}) \\
&= b_{ij} b_{mk} + O(N^{-1}).
\end{aligned}$$

Pairwise independence does not imply independence, however. But, by an analogous argument to that in the proof of Theorem 3.2, we also obtain the 'independence' of multiple lines up to an error $O(1/N)$. As we have seen in Proposition 3.28, the lineages also reach a stationary distribution $\gamma^{(N)} = (\gamma_i + O_i(1/N))$. Thus the configurations of the ancestors of the sample follow a multinomial distribution up to a small error:

$$\pi_n^{(N)}(x) = \frac{n!}{x_1! \dots x_{K^*}!} \gamma_1^{x_1} \dots \gamma_{K^*}^{x_{K^*}} + O(N^{-1}). \qquad \square$$

So far we have been interested in the stationary behaviour of the $n$th level con-

figuration process, $X^{(n)}(t)$. This is the Markov chain that describes the evolution of the family size configuration of the $n$ surviving ancestors of the sample. Next, we turn our attention to the process of changing levels. First, we estimate the probability of $X(t)$ stepping down one level ($n \to n-1$), that is, to single coalescent events.

**Definition 3.33.** *Let $h_N(x) = h_N(x_1, x_2, \ldots x_{K^*})$ denote the probability that a single pair of ancestors coalesces in the parent generation if the current configuration $X^{(n)}(t) = x$ and $\sum_{j=1}^{K^*} x_j = n$.*

Recalling that $R_n$ stands for the set of the $n$th level configurations, we point out that the function $h_N : R_n \to [0,1]$ also depends on the level $n$. But, for the sake of clarity, we suppress this dependency in our notation.

**Proposition 3.34.** *The function $h_N : R_n \to [0,1]$ may be written as $h_N(x) = N^{-1}H(x) + o(N^{-1})$, where*

$$H(x) = \sum_{k=1}^{\infty} \frac{k-1}{p_k} \left( \sum_{i=1}^{K^*} \binom{x_i}{2} b_{ik}^2 + \sum_{i=1}^{K^*} \sum_{j<i}^{K^*} x_i x_j b_{ik} b_{jk} \right),$$

*for all $x = (x_1, x_2, \ldots x_{K^*}) \in R_n$. Note that the function $H$ does not depend on the total population size. For $X_N^{(n)} = (X_1, X_2, \ldots X_{K^*})$, we also have that*

$$\mathrm{E}\, H(X_1, X_2, \ldots X_{K^*}) = \binom{n}{2} \sum_{k=1}^{\infty} \frac{k-1}{p_k} \gamma_k^2 \approx \binom{n}{2} (\mathrm{E}\, S_t - 1).$$

*Proof.* Let $x$ be any configuration vector of the $n$th level process. Then, using the

Law of Total Probability, we obtain that

$$
\begin{aligned}
h_N(x) = \sum_{k=1}^{N} & \mathbf{P}\{\text{coalescence at next time step} \mid \text{coalescing in a family of size } k\} \\
& \mathbf{P}\{\text{a pair of lineages migrate to the same state } k | X^{(n)}(t) = x\} \\
&= \sum_{k=1}^{N} \frac{k-1}{Np_k^{(N)} - 1} \left( \sum_{i=1}^{K^*} \binom{x_i}{2} b_{ik}^{(N)} b_{ik}^{(N)} + \sum_{i=1}^{K^*} \sum_{j<i}^{K^*} x_i x_j b_{ik}^{(N)} b_{jk}^{(N)} \right) \\
&= \sum_{k=1}^{\infty} \frac{k-1}{Np_k(t)} \left( \sum_{i=1}^{K^*} \binom{x_i}{2} b_{ik} b_{ik} + \sum_{i=1}^{K^*} \sum_{j<i}^{K^*} x_i x_j b_{ik} b_{jk} + O(N^{-1}) \right) \\
&= \sum_{k=1}^{\infty} \frac{k-1}{Np_k(t)} \left( \sum_{i=1}^{K^*} \binom{x_i}{2} b_{ik} b_{ik} + \sum_{i=1}^{K^*} \sum_{j<i}^{K^*} x_i x_j b_{ik} b_{jk} \right) + O(N^{-2}) \\
&= N^{-1} H(x) + O(N^{-2}).
\end{aligned}
$$

Next, we calculate the expectation of $H(X_N^{(n)})$ under the invariant measure $\pi_n^{(N)}$.

$$
\begin{aligned}
\mathrm{E}\, H(X_1, \ldots X_{K^*}) &= \sum_{x \in S_n} \pi_n^{(N)}(x_1, \ldots x_{K^*}) H(x_1, \ldots x_{K^*}) \\
&= \sum_{k=1}^{\infty} \frac{k-1}{p_k} \left( \sum_{i=1}^{K^*} b_{ik}^2 \mathrm{E} \binom{X_i}{2} + \sum_{i=1}^{K^*} \sum_{j<i}^{K^*} b_{ik} b_{jk} \mathrm{E}\, X_i X_j \right) + O(N^{-1}); \\
&\qquad \text{as } X_i \sim Bin(n, \gamma_i^{(N)}) \text{ and } X \sim Mult(n, \gamma^{(N)}) \\
&= \sum_{k=1}^{\infty} \frac{k-1}{p_k} \left( \sum_{i=1}^{K^*} b_{ik}^2 \binom{n}{2} \gamma_i^2 + b_{ik} b_{jk} \binom{n}{2} \gamma_i \gamma_j \right) + O(N^{-1} \\
&= \binom{n}{2} \sum_{k=1}^{\infty} \frac{k-1}{p_k} \sum_{i=1}^{K^*} \sum_{j=1}^{K^*} \gamma_i b_{ik} \gamma_j b_{jk} + O(N^{-1}); \\
&\qquad \text{finite series may be rearranged} \\
&= \binom{n}{2} \sum_{k=1}^{\infty} \frac{k-1}{p_k} \left( \sum_{i=1}^{K^*} \gamma_i b_{ik} \right) \left( \sum_{j=1}^{K^*} \gamma_j b_{jk} \right) + O(N^{-1}); \gamma \text{ is invariant} \\
&= \binom{n}{2} \sum_{k=1}^{\infty} \frac{k-1}{p_k} \gamma_k^2 + O(N^{-1}).
\end{aligned}
$$

Introducing the letter $c = \sum_{k=1}^{\infty} \frac{k-1}{p_k} \gamma_k^2$, we emphasise the dependence of the expectation on the level $n$ by writing

$$
\mathrm{E}\, H(X_N^{(n)}) = c \binom{n}{2} + O(N^{-1}) =: c_n + O(N^{-1}).
$$

We have not, in fact, shown that the expectation above exists. To see this, recall that the proportion of $\gamma$ and $p$ is bounded above (Remark 3.31). So there exist

$M \in \mathbf{R}^+$ such that for all $j \in \mathbf{N}^+$ we have $\gamma_j/p_j \leq M$. Now

$$\begin{aligned} c(N) &:= \sum_{j=1}^{N} \frac{j-1}{p_j}\gamma_j^2 \\ &\leq M^2 \sum_{j=1}^{N} (j-1)p_j \\ &\leq M^2(\mathrm{E}\, S - 1), \end{aligned}$$

for all $N \in \mathbf{N}^+$. Thus $c(N)$ is monotone increasing and bounded above: so converges to $c = \sum_{j=1}^{\infty} \frac{j-1}{p_j}\gamma_j^2$. □

In Lemma 3.36, we show that the quantity $c_n$ is the exponential rate with which two lineages coalesce, whenever there are $n$ surviving ancestors to the original sample. But first let us see how $c$ behaves when all the weights are close to 1.

**Remark 3.35.** Proposition 3.30 describes the invariant measure $\gamma$ of the backward migration process in the case when the reproductive weight sequence $w = (1 + \epsilon_i)$ and $\|\epsilon\|_\infty$ is small. Applying this proposition, we can now find an estimate for the scale parameter $c$ for weights close to one. Assuming that the weights are of the form $w_i = 1 + \epsilon_i$ for all $i \in \mathbf{N}^+$, and using the notation $x_i = \gamma_i/p_i$ introduced in Proposition 3.30, we obtain that

$$\begin{aligned} c &= \sum_{k=1}^{\infty} \frac{k-1}{p_k}\gamma_k^2 = \sum_{k=1}^{\infty} (k-1)p_k x_k^2 \\ &= \sum_{k=1}^{\infty} (k-1)\frac{e^{-1} + O(\|\epsilon\|_\infty^2)}{(k-1)!}(1+\epsilon_k)^2 \\ &= \sum_{k=2}^{\infty} \frac{1+2\epsilon_k}{(k-2)!}e^{-1} + O(\|\epsilon\|_\infty^2);\ \text{Taylor expansion of exp} \\ &= 1 + 2\sum_{k=2}^{\infty} \frac{\epsilon_k}{(k-2)!}e^{-1} + O(\|\epsilon\|_\infty^2). \end{aligned}$$

Lastly we observe that, as $p_k = e^{-1}/(k-1)! + O_k(\|\epsilon\|_\infty^2)$, the scale parameter $c$ is more sensitive to a small perturbation of the weights around 1 than the time stationary family size distribution $p$ is.

What is the probability that we do not see any coalescence event for $Nt$ generations? The answer is expressible in terms of the function $h$.

**Lemma 3.36** (Probability of no coalescence)**.** *Let $N \in \mathbf{N}^+$ denote the population size, $t \in \mathbf{R}^+$ and $c_n$ be the scaling factor as defined in the previous proposition, then*

$$\mathbf{P}\{\textit{no coalescence occurs for } \lfloor Nt \rfloor \textit{ generations}\} = \prod_{k=1}^{\lfloor Nt \rfloor} (1 - h_N(X^{(n)}(k))) \to e^{-c_n t},$$

*as $N$ becomes large.*

Therefore, if we measure time in units of $N$ generations and there are $n$ individuals ancestral to the original sample remaining, pairs of lineages coalesce with rate $c_n$. Note also that the limit in the lemma is a generalisation of the usual 'exponential limit', $\lim_{N\to\infty} \left(1 + \frac{a}{N}\right)^N = e^a$.

*Proof of Lemma 3.36.* The equality in the statement holds because the probability that two lineages do not coalesce in generation $k$ is $1 - h_N(X^{(n)}(k))$. So that the limit $\lim_{N\to\infty} \prod_{k=1}^{N} (1 - h_N(X^{(n)}(k)))$ exists and equals $e^{-c_n}$, it is enough to show that the following two conditions hold.

1. $\lim_{N\to\infty} \max_{1\le k\le N} h_N(X^{(n)}(k)) = 0$, and
2. $\lim_{N\to\infty} \sum_{k=1}^{N} h_N(X^{(n)}(k)) = c_n$ exists.

To convince ourselves that these conditions are sufficient, we note that the logarithm

function is continuous on $\mathbf{R}^+$ and, as $N$ becomes large, we consider

$$\begin{aligned}
\log \prod_{t=1}^{N} (1 - h_N(X^{(n)}(t))) &= \sum_{t=1}^{N} \log(1 - h_N(X^{(n)}(t))) \\
&\approx \sum_{t=1}^{N} -h_N(X_n(t)) \\
&= -\sum_{t=1}^{N} \left( \frac{H(X^{(n)}(t))}{N} + O(N^{-2}) \right) \\
&\to -c_n.
\end{aligned}$$

Condition (1) is almost surely true as we have $\sup_{t\geq 1} H(X^{(n)}(t)) \in \mathbf{R}$ a.s., as its expectation has been shown to exist in Proposition 3.34, and hence

$$\begin{aligned}
h_N(X_n(t)) &= \frac{1}{N} H(X^{(n)}(t)) + o(N^{-1}) \\
&\leq \frac{1}{N} \sup_{t\geq 1} H(X^{(n)}(t)) + o(N^{-1}) \\
&\to 0, \text{ as } N \to \infty \text{ uniformly in } t.
\end{aligned}$$

For Condition (2) we apply the Strong Law of Large Numbers for additive functionals of a Markov chain, (see Durrett (1996)).

$$\begin{aligned}
\lim_{N\to\infty} \sum_{t=1}^{N} h_N(X^{(n)}(t)) &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} H(X_n(t)) + \lim_{N\to\infty} O(N^{-1}); \text{ SLLN} \\
&= \sum_{x\in S_n} \pi_n(x) H(x) \text{ a. s.} \\
&= \mathrm{E}\, H(X^{(n)}); \text{ by previous calculation} \\
&= \binom{n}{2} \sum_{k=1}^{\infty} \frac{k-1}{p_k} \gamma_k^2 = c_n.
\end{aligned}$$

This shows that Condition (2) holds too; and the proof is complete. □

We are ready to show that the ancestral process $A_N(Nt) = |X(Nt)| := \sum_{j=1}^{K^*} X_j(Nt)$ converges weakly to the time-rescaled $n$-coalescent, $A(ct)$ in the infinite population size limit. The $n$-coalescent, sometimes also called Kingman coalescent, is a coagulation process in which any pair of surviving ancestral lines are equally likely to merge, and mergers involving more than two lineages do not occur. Definition 2.4 gives a more formal description.

In our demonstration we will closely follow the proof of a similar convergence result from (Nordborg and Krone, 2002) on page 33 for coalescent processes in which lineages migrate faster in between states than they coalesce.

**Theorem 3.37** (Nordborg and Krone (2002))**.** *Let $\Pi_N = (\Pi_N(x,y))_{x,y\in R}$ be the one step transition probability matrix for a configuration process $Y_N$ where the population size is $N$. Let $B$ and $C$ be the matrices which encode the probabilities of migration (state transitions within levels) and single coalescence events (the transition from level $n$ to level $n-1$). Then the ancestral process $A_N(\lfloor Nt \rfloor) = |Y_N(\lfloor Nt \rfloor)|$ converges weakly to the time-changed $n$-coalescent process $A(ct)$, where the rate of $n$th level single coalescence events is $c_n = \binom{n}{2}c$, provided that*

$$\Pi_N = \frac{1}{N^\alpha}B + \frac{1}{N}C + o(N^{-1}),$$

*holds for any $\alpha \in [0,1)$.*

Here the linages are supposed to reach their time-stationary distribution faster than the 2-mergers occur. This assumption holds in most cases when the types are geographically separated demes or age groups. But in our case the ancestral lines' distribution merely gets close to their invariant measure before a single coalescence event occurs.

Prior to stating the main result of the chapter, we recall some notation introduced

earlier in the text. Let $d_n$ denote the number of states of $R_n$, the set of configurations with $k$ lineages, and let $d = d_1 + \cdots + d_m$ be the total number of configurations of $R = \bigcup_{n=1}^{m} R_n$.

We also let $\Pi_N$ denote the $d \times d$ transition matrix of our Markov chain, the backward configuration process, $X_N$. $X_N(t)$ denotes the ancestral type configuration of a given sample of $m$ individuals back in time. So $|X_N(0)| = m$. Finally, the index $N$ signifies that the sample was drawn from a population of size $N$.

**Theorem 3.38.** *Provided that the family size distribution $p^{(N)} = (p_j^{(N)})$ is time-stationary, the ancestral process $A_N(\lfloor Nt \rfloor)$ started with $m$ lineages converges weakly to the rescaled $m$-coalescent $A(ct)$, as $N$ becomes large, where $c = \sum_{j=1}^{\infty} \frac{j-1}{p_j} \gamma_j^2$.*

*Proof.* By Theorem 2.12 in §4 of Ethier and Kurtz (1986), it is enough to show that

$$\lim_{N\to\infty} \sum_{y \in R_k} \Pi_N^{\lfloor Nt \rfloor}(x, y) = \sum_{j=0}^{\infty} \frac{t^j}{j!} q_{nk}^{(j)}, \tag{3.10}$$

for all levels $n$ and $k \geq 1$ and configuration $x \in R_n$. The quantity $q_{nk}^{(j)}$ is the $(n, k)$th entry of $Q^j$, the $j$th power of the infinitesimal generator of the $c$-scaled $n$-coalescent process $A(ct)$. Provided that the transition matrix of $X_N$, $\Pi_N$, is of the form

$$\Pi_N = B + \frac{1}{N} C + o(N^{-1}),$$

where the matrices $B$ and $C$ do not depend on the population size $N$, we can follow Nordborg and Krone (2002) and Hössjer (2011) in using Möhle's Lemma (Möhle, 1998) to deduce that

$$\lim_{N\to\infty} \left( B + \frac{1}{N} C + o(N^{-1}) \right)^{\lfloor Nt \rfloor} = P - I_d + e^{tG}, \tag{3.11}$$

where $G = PCP$ is the limiting infinitesimal generator of $X_N$ run on the time scale $N$. The matrix $P$ is block-diagonal, $P = \mathrm{diag}(P_1, \dots P_m)$, where $P_n$ is a $d_n \times d_n$ matrix having identical rows corresponding to the asymptotic distribution of $n$ lineages. In other words, at any level $n \leq m$ for all configurations $y \in R_n$, we have

$$P(x, y) = \lim_{t \to \infty} B^t(x, y) = \frac{n!}{y_1! \dots y_K!} \prod_{i=1}^{K^*} \gamma_i^{y_i}.$$

In our case the backward migration matrix $B^{(N)} = B + N^{-1}\bar{B} + O(N^{-2})$ has a term of $O(1/N)$ hence, in the above representation of $\Pi_N$, the matrix $C$ does not only contain transitions due to coalescence but also the residual term $\bar{B}$ from the backward migration matrix $B^{(N)}$. Thus we can use the matrix

$$C = \bar{B} + C_2,$$

$C_2$ denoting the matrix encoding state transitions due to coalescence of two lineages. Here we observe that the probability of an $l$-merger (a merger unifying $l$ lines) lies in the interval $[\min_{1 \leq j \leq K}(r_j/N)^{l-1}, \max_{1 \leq j \leq K}(r_j/N)^{l-1}]$: they are all $O(1/N^{l-1})$ rare events. Hence the state transitions due to multiple mergers are not contained in the $N^{-1}C$ term of our representation of $\Pi_N$, but in the $o(1/N)$ term.

In order to show that $c = \sum_{j=1}^{\infty} \frac{j-1}{p_j} \gamma_j^2$ is the correct scalar factor in the time scale (see Lemma 3.36), we must verify that

$$\sum_{y \in R_{n-1}} C(x, y) = H(x), \tag{3.12}$$

for all levels $2 \leq n \leq m$ and configurations $x \in S_n$, and also that

$$C(x, y) = 0,$$

for $x \in S_n, y \in S_k$ and $1 \leq k \leq n-2 \leq m-2$.

We show the first statement by noting that the migration of lineages in between family size classes cannot lead out of level $n$, hence

$$\sum_{y \in R_{n-1}} B^{(N)}(x, y) = 0,$$

for all configuration $x \in R_n$. Hence we also have

$$\sum_{y \in R_{n-1}} \bar{B}(x, y) = 0,$$

too, because $B^{(N)} = B + N^{-1}\bar{B} + O(N^{-2})$. Hence, expanding the matrix $C$, we obtain that

$$\begin{aligned}
\sum_{y \in R_{n-1}} C(x, y) &= \sum_{y \in R_{n-1}} (\bar{B}(x, y) + C_2(x, y)) \\
&= \sum_{y \in R_{n-1}} C_2(x, y); \text{ by Proposition 3.34} \\
&= H(x),
\end{aligned}$$

for all $x \in R_n$. We verify the second condition by observing that if the configurations $x \in R_n$ and $y \in R_k$ with $1 \leq k \leq n-2 \leq m-2$, then $C_2(x, y) = 0$ because a 2-merger can take us only one level lower; and also $\bar{B}(x, y) = 0$ since the matrix $B^{(N)} = B + N^{-1}\bar{B} + O(N^{-2})$ is block-diagonal: $B^{(N)}$ contains the $n$-level backward migration matrices $B_n^{(N)}$ on its diagonal.

Now from Equations (3.11) and (3.12), we obtain that

$$\sum_{y \in R_n} G^j(x, y) = q_{nk}^{(j)},$$

for all $j \in \mathbf{N}$, all levels $1 \leq k \leq n \leq m$ and configurations $x \in S_n$. Following the proof of Theorem 3.37 in Nordborg and Krone (2002), we deduce (3.10); and the proof is complete. □

**Remark 3.39** (Slow component of migration)**.** Were we not only interested in the evolution of the number of surviving lineages, we could consider the limiting infinitesimal generator of the configuration process $X_N$ which has been obtained above to be $G = PCP = P\bar{B}P \oplus PC_2P$, which is a direct sum of two generators. The familiar second one describes the transitions in between levels (vertical state transitions), while $P\bar{B}P$ encodes the migration of lineages in between family size classes (horizontal transitions). Thus the surviving lineages migrate between family size classes even on the $N$ time scale.

**Remark 3.40** (Infinite but bounded weight sequences)**.** From our argument above for estimating the chance of coalescence events involving more than two ancestral lines, it is clear that as long as the relative weights $(r_i)$ are bounded multiple mergers occur with probability $O(1/N^2)$. This suggests that the coalescent dual of our model, time measured in units of population size, with an infinite but bounded weight sequence $(w_i)$ may also weakly converge to a constant scaled $n$-coalescent. (Of course, if the process is infinite dimensional, then the matrices become operators, which require a more careful treatment.)

Although our convergence result only holds when the weights are chosen so that the family size distribution becomes time invariant, these two remarks illustrate that

the CTF extension of the Wright-Fisher model is mathematically appealing.

## 3.5 Conclusion

Cultural transmission of fitness, the phenomenon that an individual's fertility is influenced by the number of their siblings has fascinated geneticists for over a century. While Karl Pearson assumed that children inherit their father's own family size, recent studies suggests that the influence is more subtle, (we refer back to §1.3 for details). Sibert, Austerlitz and Heyer contributed to the subject by conducting computer simulations (Austerlitz and Heyer, 1998; Sibert et al., 2002). The generated data demonstrates that CTF, under their models, may not only considerably shorten the time to the most recent common ancestor to a randomly picked pair from the population, but also, it skews the shape of the trees in the genealogy. The fact that the trees are unbalanced, however, implies that the infinite population size limit of the coalescent dual of their models is not the $n$-coalescent, (see Definition 2.4).

In most populations there exists an optimal family size which confers the greatest benefit to the family members. The existence of such an optimal family size implies two facts. First, as Austerlitz and Heyer (1998) confirm, the family size distribution is usually flat around the most advantageous size. Second, individuals living in very large families should have fewer offspring on average, relative to the rest of the population. These are the reasons that we decided to retain sufficient control over the family size distribution by assigning a weight for each family size class. The only requirement in our constant population size model (Definition 2.7) on the weight sequence $w$ is that it has a constant tail. This restriction ensures that, in the infinite population size limit, the family size distribution in the forward-in-time model becomes deterministic. Looking back to the calculations in §3.1, it seems plausible that the model is well-behaved even for any bounded weight sequence – it

need not have a constant tail. At any rate, it is biologically not too restrictive to assume that two individuals brought up in very large families have the same chance to have children, irrespective of their actual family sizes.

Studying this simplified model adds more colour to the picture of CTF drawn by Sibert, Austerlitz and Heyer. Our main result that, measuring time in units of population size, the ancestral process associated with a sample of size $n$ taken from the population, under our model, converge weakly to a constant scaled $n$-coalescent, as the total population size increases, is important. It is significant because the $n$-coalescent is the limit coalescent dual of some of the most widely accepted forward-in-time models such as the Wright-Fisher and the Moran model, which is a birth-death point process. The $n$-coalescent is also the limit dual for a large class of the Cannings models, where parents have an exchangeable offspring distribution. (The Moran and Cannings models were not introduced before; we refer to any standard text book on the subject, such as Etheridge (2011).) The fact that cultural transmission of fitness, choosing an appropriate effective coalescent population size, has a place in standard coalescent theory, should therefore come as a relief to many practitioners of population genetics. Of course, there is nothing new here: geneticists have been always using the $n$-coalescent on human populations with trans-generational correlation in the family sizes. Our result justifies their long-standing practice.

In the next chapter we study the model with, the biologically less relevant, strictly monotone increasing weight sequences of the form $w = (j^\alpha)$, $\alpha > 0$. There we find that multiple mergers of lineages occur in the dual process.

We did not find evidence that larger trans-generational correlations would signal faster evolution. When the weights are close to each other, we showed that the time stationary family size distribution $p$ and the invariant measure $\gamma$, of the backward

migration process of ancestral lineages in between family size classes are almost the same. In this special case, the time-scale $c = \sum_{j=1}^{\infty} \frac{j-1}{p_j}\gamma_j^2$ is approximately just the expected number of siblings of a randomly picked individual from the population. Thus larger trans-generational correlation between the family sizes does not necessarily come with faster evolution. Instead, the level of correlation may play two roles in data analysis. First, the presence of a correlation does signal that CTF is at work; second it tells apart two populations if they evolve with different sets of weights. For example, in §1.2 we saw that within a diploid population the trans-generational correlations in maternal lines and paternal lines disagree, which implies that a father's reproductive weights differ from a mother's weights. So maternal and parental lines, in this case, will necessarily evolve at different rates.

# 4 Part II: Computer simulations

In Part I we studied the CTF extension of the Wright-Fisher model, introduced in Definition 2.7, with reproductive weight sequences of constant tail. We saw that boundedness of the weights was key in showing that the rescaled ancestral process weakly converges to the $n$-coalescent. In this chapter, we conduct computer simulations to study the case of strictly increasing weight sequences of the form $w_j = j^\alpha$, for $\alpha > 0$. Here, unlike in Part I, we will observe multiple mergers occurring in the limit coalescent, just as Sibert et al. (2002) originally predicted.

Our results are based on numerical data obtained by computer simulations. We simulated populations of different sizes $N$ forward in time ($12 \leq N \leq 1000$), evolving according to our model with monotone strictly increasing weights of the form $(j^\alpha)$. In our experiments the value of the parameter $\alpha$ varies in the range $[0, 5]$. Unlike in the previous sections, in the infinite population size limit we possibly have infinitely many family size classes. Statistics that are thought to characterise the data well were then taken. These include family sizes, numbers of reproducing individuals, and times to the most recent common ancestors of samples of different sizes.

In order to speed up the simulation, the Matlab code was designed to allow for distributed parallel computation. The process was distributed over computers according to population size, where it ran concurrently on the processors split by the weight parameter $\alpha$. Redeeming a for-cycle on vectors by a matrix multiplication (a technique often employed in Matlab), the simulation was further accelerated by around 35%. In order to compute the time to the MCRA of all pairs in each generation efficiently, we devised a function following the principles of dynamic programming. Having calculated the time to the MRCA of each pair in a given generation, we can easily work out the times to the MRCA in the next generation.

If the pair $(i, j)$ has the same parent then their time to the MRCA is 1; otherwise it is the the time to the MRCA of individual $i$'s and $j$'s parents + 1. The program's building block functions, including this one, are collected in Appendix C.

Our aim is to learn more about the infinite population limit of the model in its time stationary state. In the Wright-Fisher model (case $\alpha = 0$), the process' dynamics is known to settle down in $4N$ generations; and we observed that the statistics' distributions just become more condensed as $\alpha$ increases. Thus, simulating $4N$ generations takes us into the stationary process for all $\alpha \geq 0$.

Two experiments were performed. In the first experiment, we ran the process $K$ times for each value of $N$ and $\alpha$ (see Table 3). In order to reach the time stationary distribution, we started with a population of $N$ individuals and simulated the process for $4N$ generations. In W-F the entire population has a common ancestor in order $2N$ generations and then that progenitor is sampled from a population with a common ancestor $2N$ generations before that. Of course, this does not take variance into account, but in our experiments we saw that $\alpha > 0$ just shortens the time to the MRCA. After time stationary distribution have been reached, the parent vectors were recorded for the next $T$ generations. The $i$th entry of the parent vector belonging to a given generation is the index of the $i$th individual's parent in the previous generation.

Table 3: Parameters of the first experiment

| Population size $N$ | Number of generations $T$ | Number of runs $K$ |
|---:|---:|---:|
| 12 | 1000 | 8000 |
| 25 | 1000 | 4000 |
| 50 | 1000 | 2000 |
| 100 | 1000 | 1000 |
| 200 | 1000 | $2 \times 500$ |
| 500 | 3200 | $2 \times 100$ |
| 1000 | 6000 | 23 |

As the vectors were kept in memory, the parameters $N$, $T$ and $K$ had to be chosen such that $O(KTN \log N)$ fitted in the the computers' memory. (Each entry of the $K \times T \times N$ matrix is represented by $O(\log N)$ bits in the memory.) Also, in order to avoid missing out on temporal patterns on the scale of population size, the time span $T$ had to be larger than a few multiples of $N$.

In the second experiment, we did not record the parent-offspring relationships for all generations. Instead, having simulated the process until a single allele swept out all others, the program worked out the number of reproducing individuals in the last two generations, the time to the MRCA of a randomly drawn sample of size 2, $\lfloor N/50 \rfloor$, $\lfloor \sqrt{N} \rfloor$ and $N$ from the last generation. In addition, a tree balance statistic was also computed for the last three of these samples. The statistic used here is explained in §4.3 – it is only defined for samples of size greater than 2. Having written these statistics into file, we repeated this procedure $K$ times, always beginning with the parent vector of the last generation. Table 4 shows the corresponding values of the sample count $K$ and the population size $N$. Again, in order to enter the time stationary distribution, we discarded the first four samples of the process.

Table 4: Parameters of the second experiment

| Population size $N$ | Number of samples $K$ |
|---:|---:|
| 250 | 40000 |
| 500 | 10000 |
| 750 | 2500 |
| 1000 | 1000 |

The point of Experiment 2 was to obtain more samples exploiting the fact that, for large values of the weight parameter $\alpha$, the time to the MRCA of the entire population is much shorter than $N$.

We refrained from taking each generation as a sample for our statistics in Experiment 2. Instead, we used the data to approximate the number of samples needed

for obtaining sharp statistics. The estimation was based on the mean time to the MRCA of two individuals. The time to the MRCA of every pair was calculated for each generation. We then checked whether drawing $K$ generations at random and taking the average resulted in the same value up to four decimal places (in the case of $N = 1000$ up to one decimal place) for repeated trials. The number of samples appearing in Table 4 was chosen to be greater than this value of $K$.

The advantage of the first approach over the second is that the extracted data is reusable, should we decide to study different statistics. This is an important point, as generating the random numbers for the simulation costs much more computer time than, say, counting family sizes or backtracking genealogies. The reason behind this is that pseudo random numbers are generated on a computer by taking the remainder after a division of large numbers, while working out statistics involves index manipulation and operations on small numbers only.

Thus, the samples in the second experiment are both well-separated in time, and roughly representative. We can safely begin the discussion of the statistics.

## 4.1 Family sizes

As an individual's chance to be chosen as a parent only depends on their family size and the weight parameter $\alpha$, knowing the behaviour of the family sizes as a function of $\alpha$ and $N$ is key to understanding the process.

**Definition 4.1** (Family size frequency)**.** *We write $f_t(k)$ for the number of families of size $k$ in generation $t$ of the process. The function $f_t$ is called the family size frequency belonging to generation $t$.*

The map $f_t$ also depends on the population size $N$, but we dropped this variable from the notation. Writing $p_k^{(N)}(t)$ for the proportion of individuals of family size $k$

in generation $t$ as in Part I (§3), we have the basic identity

$$kf_t(k) = Np_k^{(N)}(t).$$

In our simulations $f_t$ randomly fluctuates over time with a very small amplitude. It thus makes sense to study its expectation with respect to $t$. Recall that assuming a truncated model, we saw in §3, that in the infinite population size limit $f_t/N = p_k^{(N)}(t)/k$ becomes deterministic (a Law-of-Large-Numbers type result). In agreement with this intuition, plotting $\mathrm{E}[f_t(k)]$ against the population size $N$, we read off from Figure 5 that $\mathrm{E}[f_t(k)]$ is proportional to $N$.

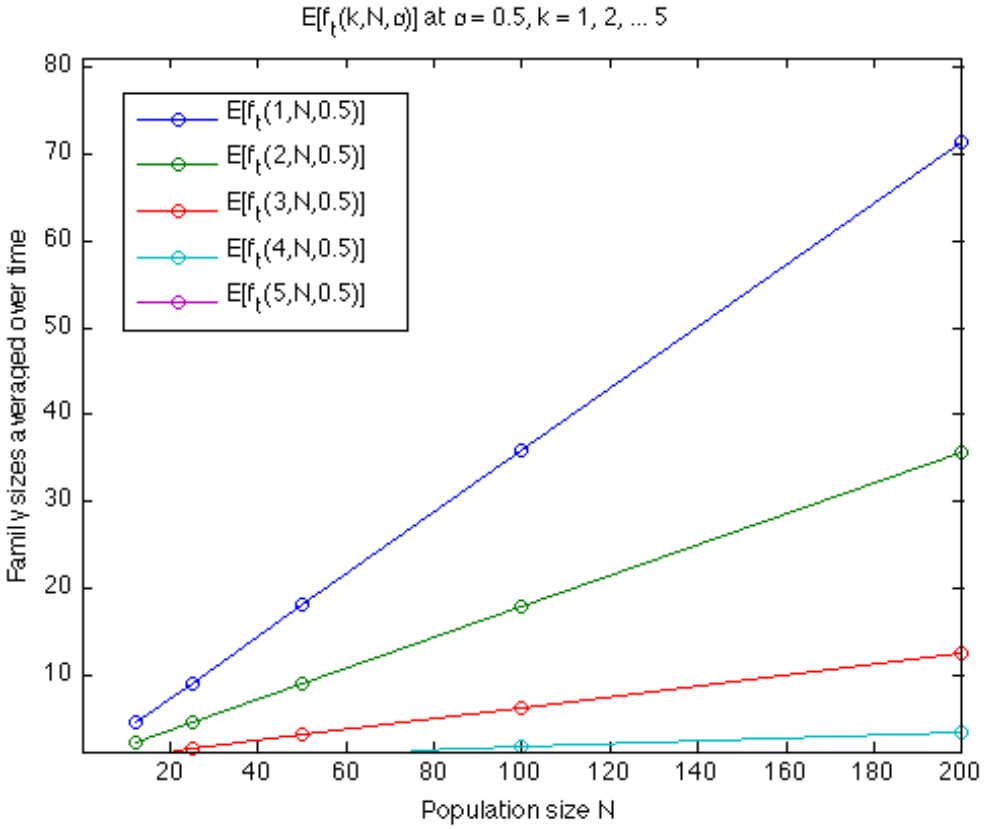



Figure 5: The mean number of families of size $k$, $\mathrm{E}[f_t(k)]$, is proportional to the population size, $N$.

Keeping $N$ constant, $\mathrm{E}[f_t(k)]$ as a function of $\alpha$ is more intricate. In the Wright-Fisher case, that is, if $\alpha = 0$, $\mathrm{E}[f_t(k)] = \frac{N}{k!}e^{-1}$. We may write $\mathrm{E}[f_t(k)]$ as a product of $N\frac{1}{k!}e^{-1}$ and a function $g(\alpha, k)$, for positive values of $\alpha$. Note that $g(0, k) = 1$ for all $k \leq N$. The function $g$ therefore can be thought of being a multiplicative deviation from the W-F case. Looking at log-log scaled graphs, we recognise that

$$\log(g(\alpha, k)) \approx \gamma(\log k)^{\frac{1}{\beta}} + \frac{1}{\delta},$$

where the parameters $\beta$, $\gamma$ and $\delta$ are all functions of $\alpha$. While the fit is perfect for values of $\alpha$ in the interval [0,1], the true $g$ increases faster than our approximation, yet slower than an exponential. As it seems in Figure 6, for the biologically more relevant family sizes ($k \leq 11$), the data points lie on the graph of the approximation even for $\alpha \in [1, 2]$.

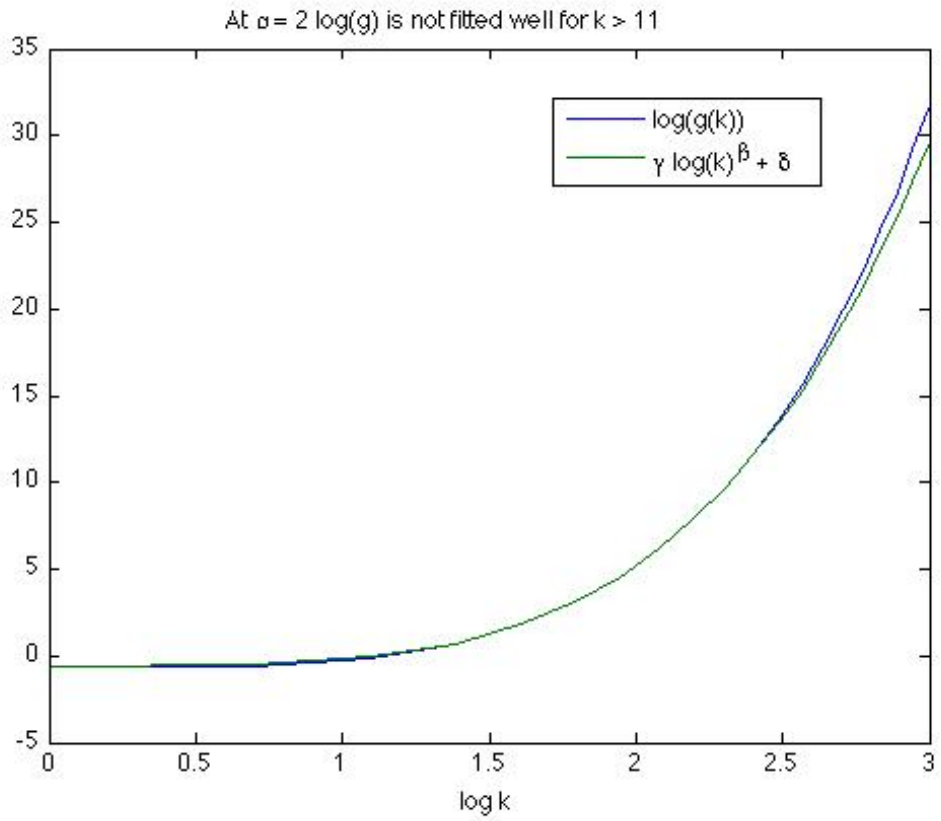



Figure 6: The approximation for the multiplicative error $g$ fits the data for smaller family sizes $k$ even at $\alpha$ on a log-log scale, but the true $g$ increases more rapidly. The fit is even better for $\alpha < 2$.

The curves of the parameters $\beta$, $\gamma$ and $\delta$ are shown in Figure 7. They all seem to be fractions of the form $\frac{a(\alpha+d)^b}{(\alpha+d)^b+c}$. The approximate values of the corresponding constants $a$, $b$ and $c$ are tabulated in Appendix B. We should note here that quite many parameters are being fitted to quite few data points.

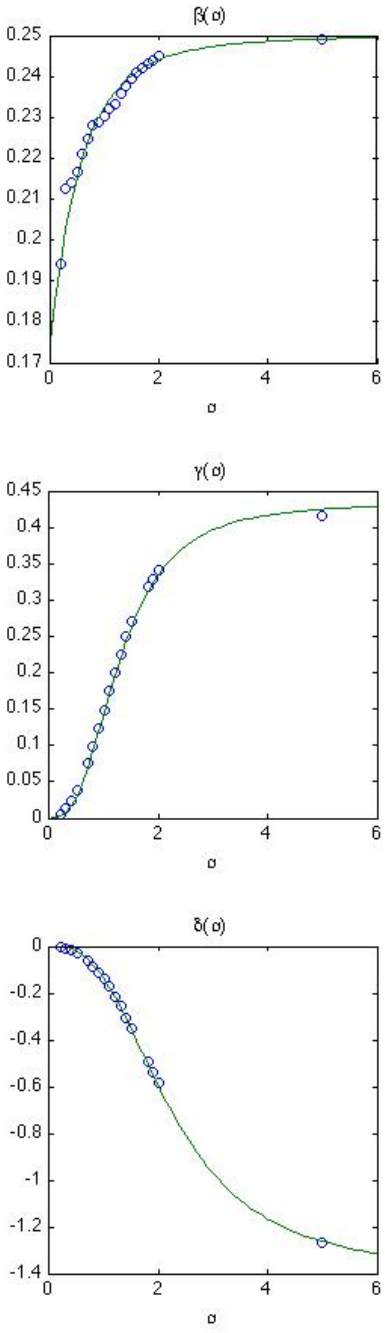



Figure 7: The parameters $\beta$, $\gamma$ and $\delta$ in the error function $g$ depend on the weight parameter $\alpha$. The blue circles are the simulated data, the green curves are the corresponding approximations.

The fractional form means a sudden change in behaviour as $\alpha$ passes $3/2$. This in turn influences the behaviour of $\mathrm{E}[f_t(k)]$. The expectation has a point of inflection at $\alpha = 3/2$. The decreasing first two family sizes are concave and turn convex at $\alpha = 3/2$, while the larger family sizes are decreasing with $\alpha$ and switch from convex to concave. As $\alpha \to \infty$, the family sizes settle down to limits shown in Appendix B. Also see Figure 8.

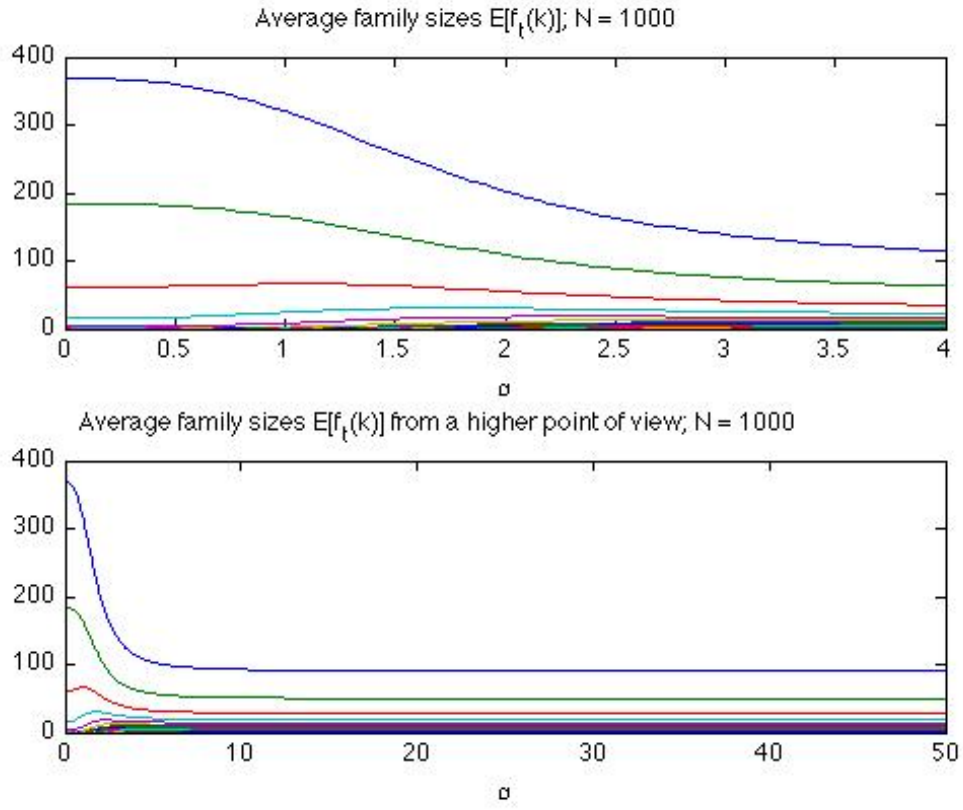



Figure 8: The blue graph shows the mean number of singleton families $\mathrm{E}[f_t(1)]$ as a function of the weight parameter $\alpha$, the green curve is of $\mathrm{E}[f_t(1)]$, and so on. Observe $\mathrm{E}[f_t(k)]$'s inflection point at $\alpha = 3/2$, and that the expectations all tend to a constant as $\alpha \to \infty$.

The significance of the phase transition at $\alpha = 3/2$ is hard to grasp, but as it turns out, it is responsible for a new development in the trans-generational correlation of average family sizes. Moreover the correlation for $\alpha \geq 2.1$ turns negative; it

is approximately $-0.52$ at $\alpha = 5$. Thus, for a given level of correlation, there correspond multiple values of the parameter $\alpha$. Sibert et al. (2002) does not point out this important circumstance, but a study using a different model, Nei and Murata (1966) predicts a maximum of 0.4 in parent-offspring correlation (of effective population size for variance), concordant with our findings. In deciding which value of $\alpha$ out of the two possibilities is relevant to a given observed correlation, tree balance comes to our aid. In the next section we define a balance statistic on ancestral trees. It turns out that the value of the tree's imbalance decides which of the two possible values of the weight parameter $\alpha$ we seek.

In order to study the mean number of reproducing individuals, $\sum_{k=1}^{N} \mathrm{E}\left[f_{t+1}(k)\right]$, we use well-separated data from the second experiment. Figure 9 shows the graphs of the average number of reproducing daughters in two successive generations. While they decrease at the same rate for $0 \leq \alpha \leq 3/2$, the two averages start to diverge at $\alpha = 3/2$. Thus the memory of the process with respect to the average number of reproducing individuals (or mean family size) is at least two-generation-deep in this regime of the weight parameter $\alpha$. More could be said about the memory, if we had recorded the numbers of the families of different sizes in the last three or four generations in the experiment.

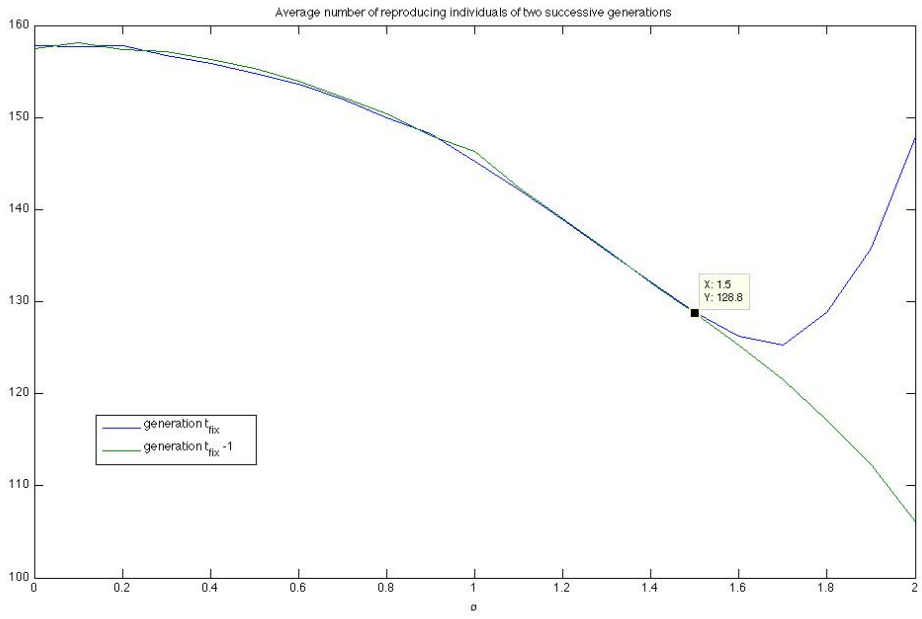



Figure 9: The blue graph shows the average number of reproducing individuals in the generation in which fixation happens. The green one is the average in the previous generation. The two curves start to diverge at $\alpha = 3/2$.

This phenomenon is not unique to this specific choice of the weight sequence $(w = (j^\alpha))$. In Example 3.9 we considered the weight sequence $w = (1, w_2, 1, 1, \dots)$. We saw that the time series of the family size class sizes converges in time for $w_2 \leq 20$ but has a limit-cycle of period two when $w_2$ is chosen to be greater than 24.

We have a cursory explanation for both the above phenomenon and the negative parent-offspring correlation for large values of $\alpha$. When the heavy weighting results in a few very large families, the next generation is sampled with equal probability from within these families, and hence the number of reproducing daughters increases again.

## 4.2 Number of reproducing daughters

The number of reproducing individuals in generation $t$ is the number of families in generation $t+1$; with the notation introduced in Definition 4.1, it is $\sum_{k=1}^{N} f_{t+1}(k)$. Hence the average family size is $N$ divided by the number of reproducing individuals. As the two statistics, the average family size and the number of reproducing individuals, are closely related and the trans-generational correlation among them produce similar graphs, we show only the correlation in the number of reproducing individuals as a function of the weight parameter $\alpha$ in Figure 10. Another reason for selecting this is that the correlation in average family sizes is less resolved. We need many more realisations.

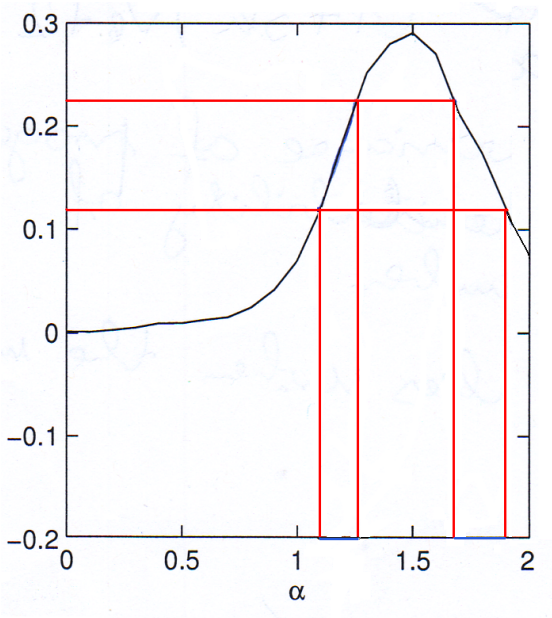



Figure 10: The mean trans-generational correlation in the number of reproducing individuals attains its maximum at $\alpha = 3/2$ for large $N$. The accentuated intervals on the $\alpha$-axis are the projections of the range in which observed correlations fall in human populations.

On the discrete domain, on which the experiment was performed, the correlation attains its maximum ($\approx 0.29$) at $\alpha = 1.4$ for $N = 25$ and 50, while the function

reaches its greatest value at $\alpha = 1.5$ when $N = 100$, 200, 500, 1000. We infer that, as $N \to \infty$, the trans-generational correlation in the number of reproducing individuals attains its maximum at $\alpha = 3/2$ and that the convergence is fast.

Figure 10 is also augmented by the observed demographic data collected from the literature. Fisher (1930) reports a family size correlation of 0.21 among women; Huestis and Maxwell (1932) reveals a correlation of 0.12; and Berent (1953) of 0.19. In the Saguenay-Lac-St-Jean region the trans-generational correlation in family sizes is 0.144 for men and 0.174 for women (Austerlitz and Heyer, 1998). In these accounts, the size of the population varies but, as we have seen, this has no significant effect on the graph: the correlation function converges fast. These data are all attained. Observe, however, that because the function is not monotonic, there belong two values of $\alpha$ for each data point in the range $[0, 0.29]$.

## 4.3 Tree topology

While tree balance is a 'horizontal' statistic on the ancestral trees, the time to the most recent common ancestor (MRCA) informs us of the height of the tree. Studying both provides insight into the effect of the weight parameter $\alpha$ on the topology of the trees. We begin by showing simulated genealogies of small populations ($N = 10$) with three different values of $\alpha$ ($\alpha = 0$, 1 and $3/2$) in Figure 11. Observe how the time to the MRCA shrinks and the ancestral trees become more and more biased as $\alpha$ increases. The intuitive reason for the decrease in the time to the MRCA is the presence of very large families. The imbalance can be explained similarly. Some branches of the genealogical trees will develop big families and some small ones; but then the descendants of large families will be again large, while the descendants of small families are likely to stay small or die out.

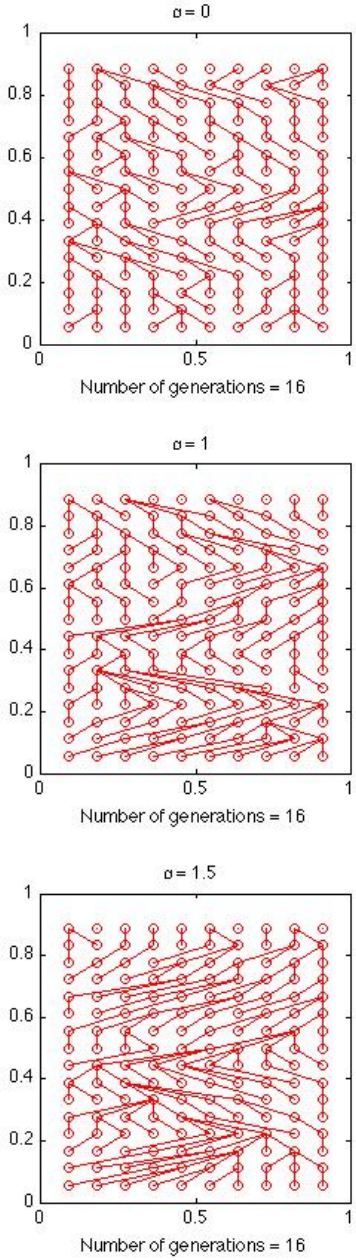



Figure 11: In these graphical representations of the time-stationary process, rows of circles signify generations of individuals, and edges represent parent-offspring relationship among them. Time runs downwards.

**Tree balance** In order to gauge the imbalance of the ancestral trees, Blum et al. (2006) employed the so called $I'$ statistic. The authors proposed $I'$ as an alternative to the parent-offspring correlation in inferring the value of the weight parameter $\alpha$ when generation by generation data is not available, see §1.1.

**Definition 4.2** (The $I'$ balance statistic)**.** *Given a rooted tree $\tau$ with $n$ leaves, $b$ leaves on its largest branch and $k$ sub-trees developing from its root, the $I$ statistic is defined by*

$$I(\tau) = \frac{b - \lfloor \frac{n}{k} \rfloor}{n - k + 1 - \lfloor \frac{n}{k} \rfloor}.$$

*The $I'$ statistic of the same tree $\tau$ equals $I(\tau)$ if $n$ is odd, and $\frac{n}{n-1} I(\tau)$ otherwise.*

The value of $I'$ describes how balanced a tree is: $I'$ is 0 when the tree is perfectly balanced – the more biased a tree is, the closer $I'$ shifts to 1. Henceforth we mean $I'(\tau)$ by the balance or imbalance of a tree $\tau$.

When the sample size $n$ is odd we have $\mathrm{E}[I] = \frac{1}{2}$ assuming that all trees are equally likely; while $\mathrm{E}[I] = \frac{n}{2} \frac{1}{n-1}$ for $n$ even. Hence, the advantage of using $I'$ over $I$ is that $\mathrm{E}[I'] = \frac{1}{2}$ for all values of the sample size $n$ in the Wright-Fisher case. The ancestral trees are not equally likely for $\alpha > 0$, however. Thus we cannot expect $I'$ to be independent of $n$ in general. We observe how the heights of the sigmoid function depends on $n$ in Figure 12. The level of the plateau increases progressively for small values of $n$; and the growth slows down as $n$ approaches the total population size $N$. Comparing the mean $I'$ statistics with varying population sizes, it is clear that $\mathrm{E}[I']$ is not a function of merely $\frac{n}{N}$. Instead, it is determined both by $n$ and $N$ separately. It would require more samples to try to find a functional form for $\mathrm{E}[I']$. (We calculated the statistic with only three different sample sizes per values of population size: $n = N/50$, $\lfloor \sqrt{N} \rfloor$ and $N$.) Nevertheless, we can measure the abruptness of the change of $\mathrm{E}[I']$ around $\alpha = 1$, by estimating $\mathrm{E}[I']$ as a fraction

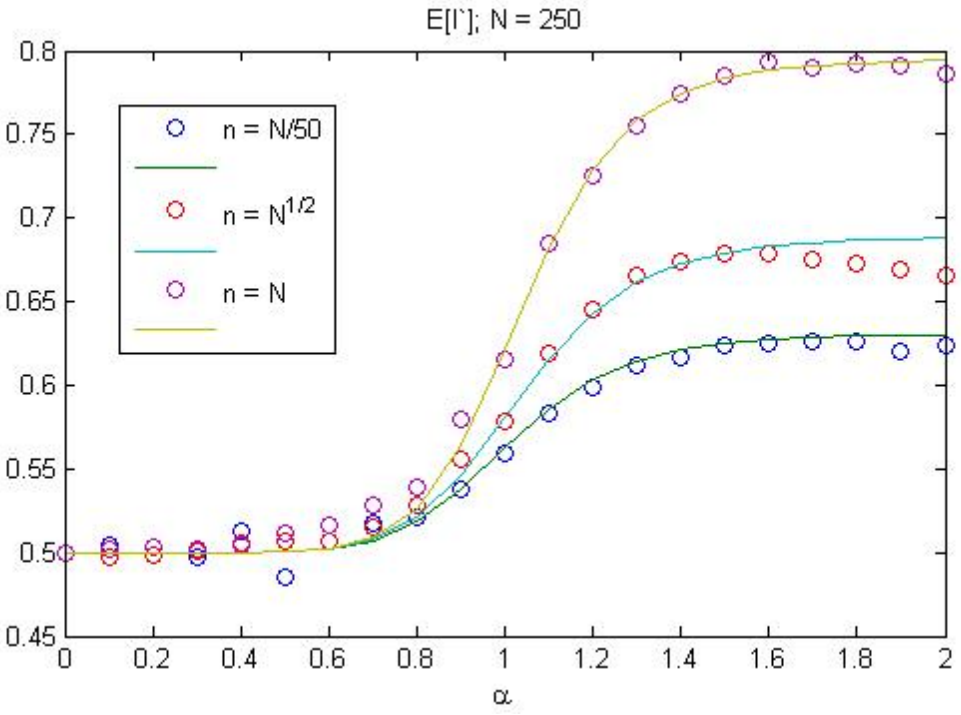



Figure 12: The mean balance $\mathrm{E}[I']$ might be estimated by a fraction for values of the weight parameter $\alpha \in [0, 2]$. Circles are simulated data, continuous graphs are approximations; the total population size, $N = 250$.

of the form $a\frac{\alpha^b}{\alpha^b+c} + \frac{1}{2}$. A few values of the parameters $a$, $b$ and $c$ are tabulated in Appendix B. Now the steepness of the function is characterised by both $a$, $b$ and $c$ because

$$\left(a\frac{\alpha^b}{\alpha^b+c} + \frac{1}{2}\right)' = a\left(\frac{\alpha^b}{\alpha^b+c}\right)' = a\frac{bc\alpha^b}{(\alpha^b+c)^2},$$

and hence the curve's gradient at $\alpha = 1$ is $\frac{abc}{(1+c)^2}$.

For $N = 250$ and $n = 5$, $b$ is approximately 7.72, while it increases to 10.14 for $N = 1000$ and $n = 50$. It might be the case that $b \to \infty$, as $N$ increases. This would mean that in the infinite population limit, $\mathrm{E}[I'] = \frac{1}{2}$ for $0 \leq \alpha < 1$, and has a jump at $\alpha = 1$. What happens after the discontinuity is difficult to tell as our approximation does not fit $\mathrm{E}[I']$ for large values of the weight parameter $\alpha$.

For $N = 250$ and $n = 5$ the 40000 simulated ancestral trees are either perfectly

balanced ($I' = 1/2$) or totally parti pris ($I' = 1$) for all values of $\alpha$. This remains the case for approximately 18000 out of the 40000 trees for $n = \lfloor 5\sqrt{10} \rfloor$, a larger sample size. Moreover, Blum et al. (2006) points out that the variance of $I'$ is so large for small values of $\alpha$ that inferring $\alpha$ from $I'$ is impossible. Fortunately, for $\alpha > 3/2$, the power statistic of the method is larger than 0.85; so the $I'$ statistic is excellent for distinguishing between the two slopes of the trans-generational correlation in average family size or in the number of reproducing individuals.

**Time to the most recent common ancestor** We write $T_n$ for the time to the most recent common ancestor (MRCA) of a sample of $n$ daughters whose genealogy is governed by our extension of W-F with the weight sequence $w = (j^\alpha)$. Plotting $T_2$ against the population size $N$ for different fixed values of $\alpha$ (see Figure 13), we observe that

$$\mathrm{E}\,[T_2] \approx N^{\gamma+\epsilon},$$

where $\gamma$ is a function of $\alpha$ but independent of $N$, and $\epsilon$ is an error term which vanishes as $N \to \infty$. That is, the time to the MRCA of a pair can be approximated as a power less than 1 of the population size.

In order to analyse the behaviour of the 'error' $\epsilon$, we may view $\frac{\log \mathrm{E}\,[T_2]}{\log N}$ as an hyperbola $\beta N^{-\delta} + \gamma$ for $\alpha > 1$, where the parameters $\beta$ and $\delta$ are functions of $\alpha$, and $\gamma$ is the horizontal asymptote. The method shows that $\epsilon \to 0$ because although $\delta(\alpha)$ is not well resolved, it is positive in the entire domain of our simulation and it presumably remains above 0 for larger values of $\alpha$ as well. Figure 14 shows an example of fitting an hyperbola to the data with $\alpha = 3/2$; and Figure 15 shows approximate values of $\gamma$. We speculate that, for $\alpha > 1$,

$$\gamma(\alpha) \approx \frac{2}{3}\alpha^{-\frac{5}{2}}.$$

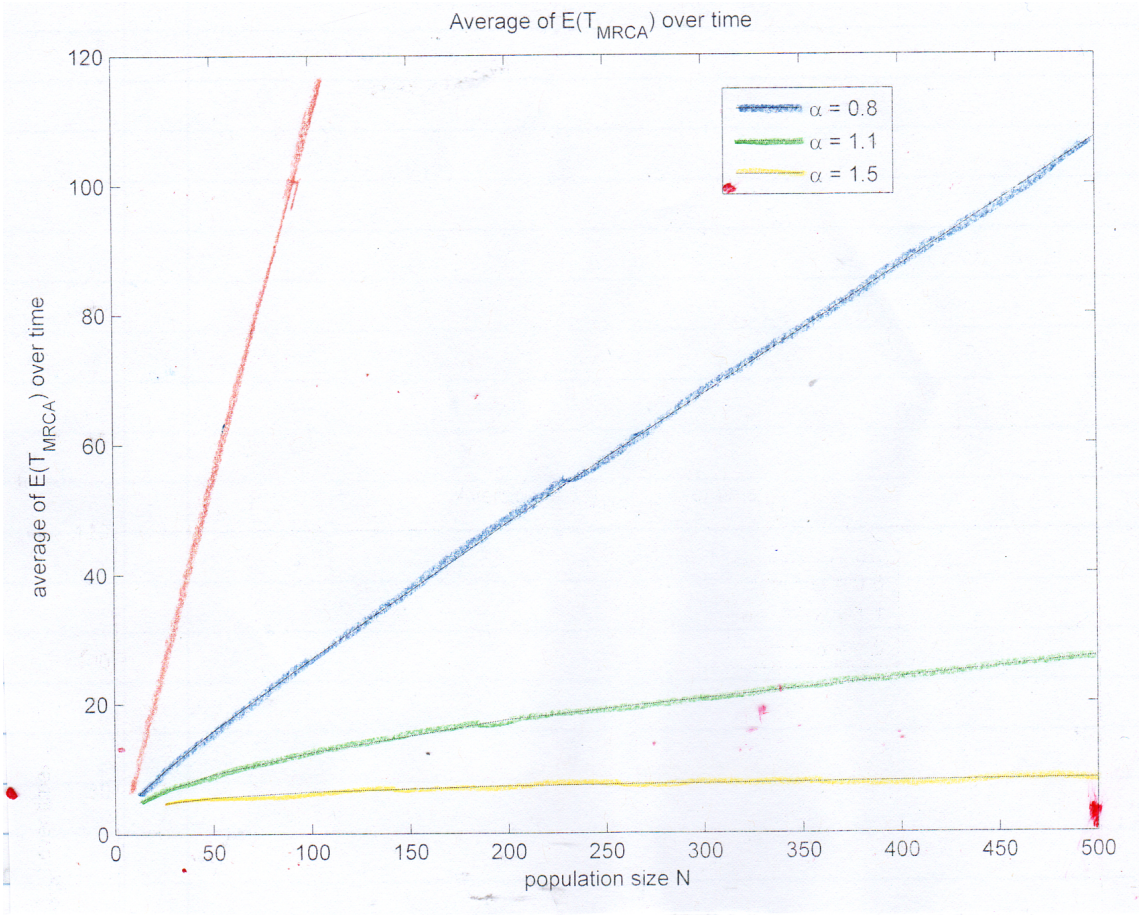



Figure 13: The mean time to the most recent common ancestor of a pair resembles a power of the population size. In the figure, the average time to the MRCA is plotted against the population size, for three fixed values of the weight parameter $\alpha$ : 0.8, 1.1 and 1.5.

When $0 \leq \alpha \leq 1$, $\epsilon$ seems to vanish, too as $N \to \infty$. In this parameter regime $\frac{\log \mathrm{E}[T_2]}{\log N}$ decreases until around $N = 100$ like a hyperbola does, but then it increases monotonically, approaching a constant $\gamma$. $\frac{\log \mathrm{E}[T_2]}{\log N}$ resembles the function $\frac{\beta}{\alpha^\delta} + \zeta \frac{\alpha^\delta}{\alpha^\delta + \eta} + \theta$, where $\gamma$ would equal $\zeta + \theta$ if the function fitted $\frac{\log \mathrm{E}[T_2]}{\log N}$ flawlessly. The data suggests that $\delta$ is an increasing function of the population size $N$.

In summary, for all $\alpha \geq 0$, $\frac{\mathrm{E}[T_2]}{N^\gamma} \to 1$ as $N \to \infty$ : measuring time in $N^\gamma$ generations is the coalescent process' natural time scale. Moreover a phase transition occurs at $\alpha = 1$.

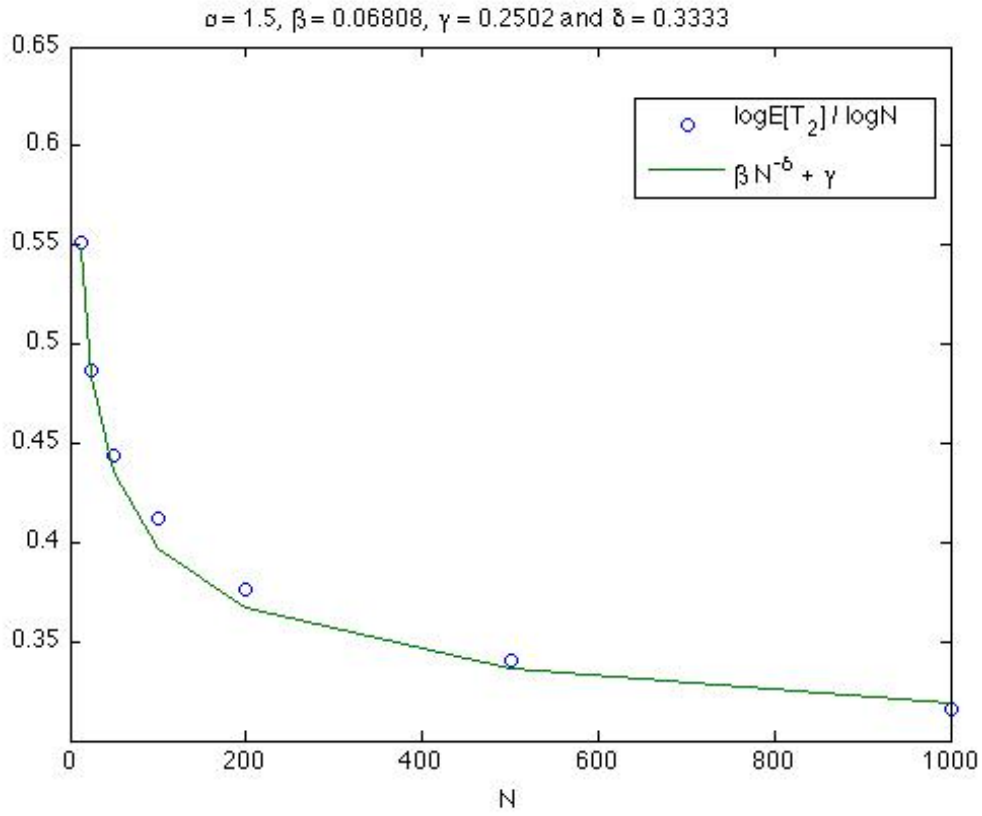



Figure 14: The quantity $\frac{\log \mathrm{E}\,[T_2]}{\log N}$ is approximated by a hyperbola of the form $\beta N^{-\delta}+\gamma$ at $\alpha = 3/2$.

We should again emphasise that, gauging the average time to the MRCA of samples of only three different sizes, obstructs inferring anything meaningful about how the size of the sample affects the statistic. We do not have the data to recover $\mathrm{E}\,[T_n]$ as a function of the sample size $n$. For comparison, we plot $\mathrm{E}\,[T_2]$, $\mathrm{E}\,[T_{\lfloor\sqrt{N}\rfloor}]$, $\mathrm{E}\,[T_N]$ and the mean time to fixation, $\mathrm{E}\,[T_{fix}]$, as functions of $\alpha$ for a fixed value of $N$ in Figure 16.

Lastly it is worth comparing these results with the literature. Campbell (1999) works with a diffusion model. He multiplies the single generation sampling variance in the fixation time's diffusion approximation by an expression which depends on the desired correlation. The author shows that, in his model, the average coalescent

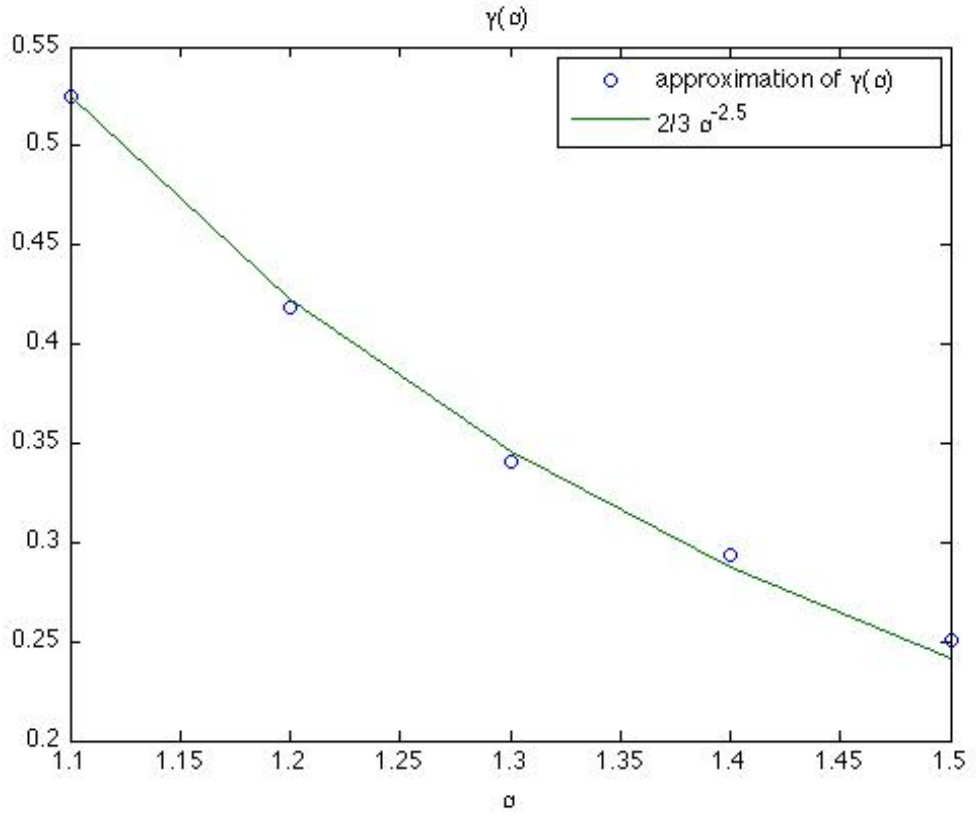



Figure 15: The power $\gamma(\alpha)$ behaves as an hyperbola for $\alpha \in [1.1, 1.5]$. The circles mark values of $\gamma(\alpha)$ calculated from the approximations. The continuous curve is $\alpha \mapsto \frac{2}{3}\alpha^{-\frac{5}{2}}$.

time equals the mean time to fixation, and that their common value is of $O(\log N)$ for all values of the model's parameter. In our case, $\mathrm{E}[T_N] < \mathrm{E}[T_{fix}]$, as it seems in Figure 16; and they are both of $O(N^\gamma)$ (with two different values of $\gamma$ ). Since $0 < \gamma < 1$, coalescence in Campbell's model is faster than in ours. Interestingly, as the weight parameter $\alpha$ increases, in our model too, $\mathrm{E}[T_N]$ and $\mathrm{E}[T_{fix}]$ become close to each other.

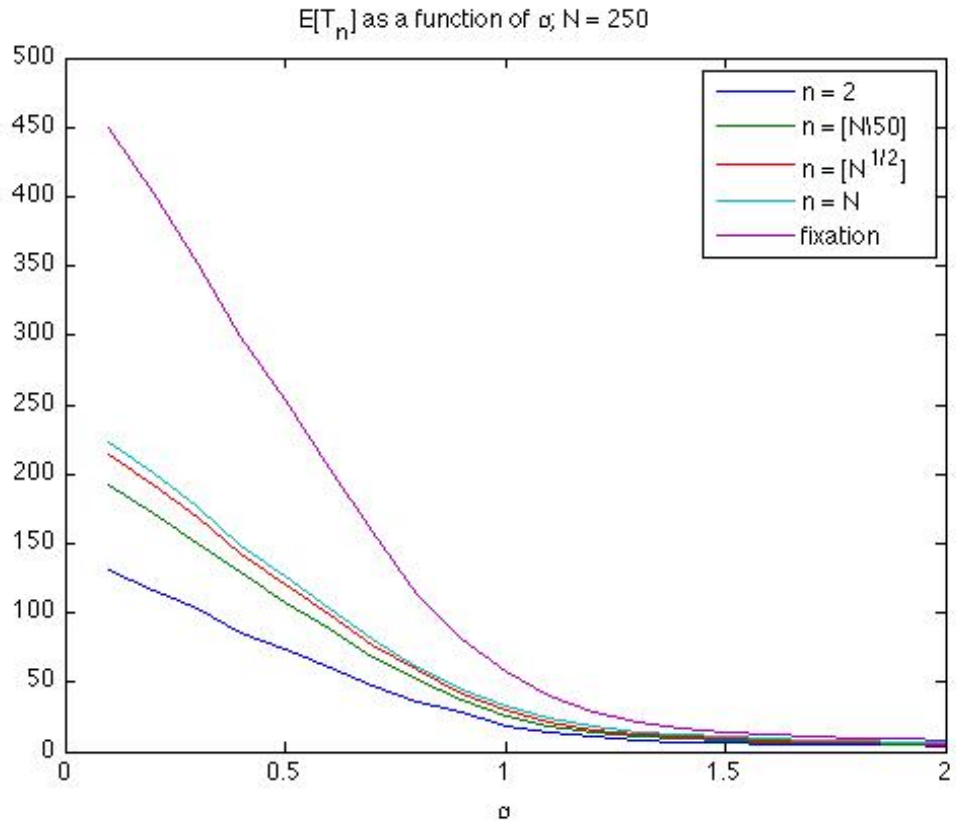



Figure 16: Mean times to the most recent common ancestor of different number of individuals, as functions of the weight parameter $\alpha$, are juxtaposed. Here the total population size $N = 250$.

## 4.4 Conclusion

In this part we performed numerics of our extended Wright-Fisher model with weight sequences of the form $w = (j^\alpha)$ for $\alpha \geq 0$. This set-up resembles Sibert, Austerlitz and Heyer's model introduced in §2.3. Fertility correlation of a given population can be estimated only by looking at ample genealogical data, and even when enough data is available such as in Iceland or the Saguenay-Lac-Saint-Jean region, Canada, the weight parameter $\alpha$ is not uniquely determined by the average parent-offspring correlation in the number of reproducing individuals. In order to select between the two possible values of $\alpha$, we can rely on the tree balance statistics $I'$ because it

proved to be informative for the biologically apposite, larger values of $\alpha$ ($1 < \alpha \leq 2$) in Blum et al. (2006). Via family sizes, CTF influences the tree topology. For relevant values of the weight parameter $\alpha$, the sigmoid function $a\frac{\alpha^b}{\alpha^b+c} + \frac{1}{2}$ explains the imbalance of the ancestral trees well. The level of the function's plateau ($a + \frac{1}{2}$) depends on both the sample size and the size of the entire population.

The trans-generational correlation in family sizes turn negative for large $\alpha$ because daughters coming from a big family which dominates the whole population have equal probability to reproduce as in W-F, so they will form smaller families themselves.

Compared to the Wright-Fisher model, our extension has a Byzantine structure which does not necessarily simplify even in the infinite population limit unless the weight sequence has a constant tail. For such weight sequences, we proved in §3.4, that the process' time-rescaled coalescent dual converges weakly to the $n$-coalescent. With the strictly monotone increasing weight sequences of the form $w = (j^\alpha)$, $\alpha \geq 0$, we see, however, that the average time to the most recent common ancestor of a pair is approximately $N^{\gamma(\alpha)}$. Thus, instead of $N$ under the Wright-Fisher model, the natural time scale of our process, with these increasing weight sequence, is $N^{\gamma(\alpha)}$. We found that $\gamma(\alpha) \approx \frac{2}{3}\alpha^{-\frac{5}{2}}$, for $\alpha > 1$. The picture becomes more complex when $0 \leq \alpha \leq 1$; in this regime $\gamma(\alpha) \approx \frac{\beta}{\alpha^\delta} + \zeta\frac{\alpha^\delta}{\alpha^\delta+\eta} + \theta$. The time scale thus goes through a phase transition at the point $\alpha = 1$. The mean balance of the ancestral tree $\mathrm{E}\, I'(\tau)$ also has a phase transition. The statistic jumps onto a higher level at $\alpha = 1$ from its constant $1/2$ value for $0 \leq \alpha < 1$.

# 5 Part III: A multi-type branching process for cultural transmission of fitness

In Part I and II we studied a model for cultural transmission of fitness where the population size was assumed to remain constant over time. Fixing the size of the population meant that individuals in any given generation did not reproduce independently from each other. There, individuals competed against each other because the total number of children – the size of the next generation – was fixed. As we pointed out in §2, while competition to reproduce is usually not negligible in human populations, in some small and rapidly expanding communities the competition can be ignored: among the Hutterites, for example, marriages are matched by the colony leaders, the Ministers (Hostetler, 1997). Thus the population dynamics in these cases could be better modelled with a branching process where all individuals reproduce independently from each other, and the total population size is allowed to vary from generation to generation. In this part we calculate the drift and diffusion coefficients of the diffusion approximation of a near-critical branching process model for CTF.

As in Part I and II, we differentiate individuals by their family size. An individual of family size $j$ receives reproductive weight $w_j$: the distribution of its offspring depends on $w_j$. Thus the number of siblings influences the number of children.

First we record the definition of a multi-type Galton-Watson process.

**Definition 5.1** (General multi-type Galton-Watson process with $K$ types)**.** *Let $Z_i(t)$ denote the number of $i$-type individuals alive in generation $t$. The population, described by the vector $Z(t) = (Z_1(t), \ldots Z_K(t))^T$, evolves as follows: each individual alive at generation $t$ produces a random assortment of offspring encoded by a vector $l = (l_1, \ldots l_K) \in \mathbf{N}^K$ independently of the rest. Here $l_j$ denotes the number of the*

*individual's $j$-type offspring. After reproduction the individuals die; their offspring form generation $t+1$. An $i$-type individual's offspring vector is $l = (l_1, \ldots l_K)$ with probability $p(e_i, l)$, where $(e_i)_{1 \leq j \leq K}$ is the canonical basis of the vector space $\mathbf{R}^K$.*

In order to apply the general theory to our branching CTF model, we specify the probabilities $p(e_i, l)$. As in CTF an individual's children are all of the same type, the offspring vector $l$ simplifies. An individual of family size $i$ has $j$ offspring of type $j$ with probability $p(e_i, je_j) := p_{ij} = p_j(w_i)$, where $(p_j(w))$ is a probability mass function of mean $w$ and finite variance. We require furthermore that for each $j \in \mathbf{N}^+$, $p_j$ is twice differentiable with respect to the parameter $w$. Thus, the distribution $p$ is parametrised by its mean, and satisfies

$$\sum_{j=1}^{\infty} j p_j(w) = w,$$
$$\sum_{j=1}^{\infty} j p_j'(w) = w' = 1 \text{ and}$$
$$\sum_{j=1}^{\infty} j p_j''(w) = w'' = 0.$$

Now we describe the model in terms of the distribution $p$.

**Definition 5.2** (Multi-type Galton-Watson process for CTF)**.** *Consider a population of individuals classified by their family size evolving throughout the generations. Let $Z_i(t)$ denote the number of individuals of family size $i$ in generation $t$. Write $\xi_{k,i}(t)$ for the number of offspring of the $k$th individual of family size $i$ in generation $t$. Then the probability mass function of $\xi_{k,i}(t)$ is $p(w_i) = (p_j(w_i))$ for all $1 \leq k \leq Z_i(t)$ and $i$, $t \in \mathbf{N}$. Also, $\{\xi_{k,i}(t)\}$ is an independent family of random variables. We say that the stochastic process $Z(t) = (Z_i(t))_{i \in \mathbf{N}^+}$ is a CTF branching process with weight sequence $w = (w_i)$.*

Thus the variables $\xi_{k,i}$ have a different distribution from the vector $l$. $\xi_{i,k}$ is the number of the $k$th $i$-type individual's children, while $l$ is the offspring vector of *any* individual.

**Remark 5.3.** The total population size in generation $t+1$ is

$$|Z(t+1)| = \sum_{i=1}^{\infty} Z_i(t+1) = \sum_{i=1}^{\infty} \sum_{k=1}^{Z_i(t)} \xi_{k,i}(t).$$

However, we allow ourselves to sometimes refer to the vector $Z(t)$ as the population size in generation $t$, too.

**Remark 5.4.** Our choice of the distribution for the number of offspring of an individual with a given family size may seem too arbitrary. Recall that, in our extension of the Wright-Fisher model in Part I, the offspring number of an individual of family size $j$ has distribution $Bin(N, r_j/N)$, (see Definition 2.7). As the total population size $N$ becomes large this distribution approaches $Pois(r_j)$. Thus, choosing Poisson distributions for the offspring numbers in this model, makes the two models somewhat comparable when the populations are numerous. We need to be aware of the fact, however, that in our first model $(r_j)$, the infinitesimal relative weights are random variables themselves: they fluctuate from generation to generation. We should also remind ourselves that this large population assumption cannot always hold in branching CTF. Population sizes randomly fluctuate in branching processes. So much so that they can even entirely die out.

Continuing the comparison with the constant population size model, if we set the weight sequence $w$ in our branching process to equal $(1 + \epsilon_i/N)$ with $\epsilon \in l^{\infty}$ and $N \in \mathbf{N}$ large, then in §5.3 we shall see that the family size distribution of the limit process $\lim_{N\to\infty} \frac{1}{N} Z(\lfloor Nt \rfloor)$ is $\left(\frac{e^{-1}}{(j-1)!}\right)$ for the Poisson offspring distributions, and $(jp_j(1))$ in general. This agrees with our result for the constant population size

model counterpart in Proposition 3.11. Nevertheless, we saw in §1.3 that human family size distribution is typically not Poisson, but tends to flatten out around the economical or societal optimal size of the family.

As in Part I, we start by considering truncated weight sequences; that is, we assume that there exists some $K \in \mathbf{N}^+$ such that $w_i = w_K$ for all $i \geq K$. Here too, we introduce a special type $K^*$ of individuals whose family size is so large that they can be taken to have the same reproductive fitness. We say that an individual is of type $K^*$ if its family size is at least $K$. These individuals all have $\xi_{k,K} \sim p(w_K)$ many children. This construction reduces our countably many types Galton-Watson process into a simple multi-type branching process with $K$ types; and hence we will be able to apply Budhiraja and Reinhold (2010) to deduce the law of $\lim_{N\to\infty} N^{-1}Z(\lfloor Nt \rfloor)$.

The purpose of §5.1 is to introduce basic techniques of analysis for multi-type branching processes. We shall see there that truncation is counter productive: for large values of $K$, it becomes difficult to recover a key quantity, the greatest eigenvalue (in modulus) of the mean matrix associated with the branching process. As it turns out, the analysis of the infinite-dimensional process is simpler. This will be carried out in §5.2. Finally, we will improve on §5.2 in §5.4 by allowing strictly increasing unbounded weight sequences.

By construction, $Z(t)$ is a Markov chain with a countable state space. However, much more is known of processes of its type: there is a rich theory of Galton-Watson processes developed over the last century. We start applying these results by calculating the mean and variance of $Z(t)$ and describing the long time behaviour of the process. In §5.3, we study a diffusion approximation of the model equipped with a special class of weight sequences. The populations that we aim to model have survived to present time, we therefore need to work with a branching process

conditioned on non-extinction. Budhiraja and Reinhold (2010) showed that the sequence of conditioned Galton-Watson processes converges weakly to the diffusion limit conditioned similarly. Applying their result, we will be able to study the proper limiting CTF process.

## 5.1 The mean and the variance matrices, the Perron root and the long time behaviour of the process

Much of the population dynamics in a Galton-Watson process is determined by the mean and the variance of the offspring number. In the multi-type case, the respective notions are the mean and variance matrices associated with the branching process. A good general introduction to multitype G-W processes (with biological applications) is Kimmel and Axelrod (2002).

**Definition 5.5** (Mean and variance matrices)**.** *Define the $K \times K$ mean matrix $M = (m_{ij})$ and the variance matrix $S = (\sigma_{ij})$ by writing $m_{ij} = \mathrm{E}^{e_i} Z_j(1)$ and $\sigma_{ij} = \mathrm{var}^{e_i} Z_j(1)$.*

Here $\mathrm{E}^{e_i}$ means expectation with respect to the law of $Z$ started with one individual of family size $i$. Thus $m_{ij}$ is the expected number of $j$-type daughters of an $i$-type mother, and $\sigma_{ij}$ is the variance of the $j$-type offspring of an $i$-type individual.

In this section, we illustrate the spectral theory of multi-type G-W processes by calculating the mean matrix, its principal eigenvalue with the corresponding left and right eigenvectors, and the variance matrix in the case of Poisson offspring distributions. We will return to the general case in §5.2.

**Definition 5.6.** *In a $K$-truncated Galton-Watson process for CTF an $i$-type individual's offspring vector is $l = (l_k)$, where $l_k = je_j$ with probability $p(e_i, l) = \frac{w_i^j}{j!} e^{-w_i}$ for all $1 \leq i < K$ and $l = je_j$ with probability $p(e_i, l) = \frac{w_K^j}{j!} e^{-w_K}$ for $i \geq K$.*

Calculating the entries of the mean matrix $M$, for $1 \leq i, j < K$, we obtain that

$$\begin{aligned} m_{ij} &= \mathrm{E}^{\,e_i} Z_j(1) = \mathrm{E}\,[\text{number of } j\text{-type offspring of an } i\text{-type individual}] \\ &= \sum_{k=0}^{\infty} k\mathbf{P}\{\text{an individual of family size } i \text{ has } k \text{ } j\text{-type offspring}\} \\ &= j\mathbf{P}\{\text{an individual of family size } i \text{ has } j \text{ offspring}\} \\ &= j\frac{w_i^j}{j!}e^{-w_i} = \frac{w_i^j}{(j-1)!}e^{-w_i}. \end{aligned}$$

Similarly, for $j < K$, we observe that

$$\begin{aligned} m_{K^*j} &= \mathrm{E}\,[\text{number of } j\text{-type offspring of a } K^*\text{-type individual}] \\ &= \frac{w_K^j}{(j-1)!}e^{-w_K}. \end{aligned}$$

Finally, for $1 \leq i \leq K^*$,

$$\begin{aligned} m_{iK^*} &= \mathrm{E}\,[\text{number of } K^*\text{-type offspring of an } i\text{-type individual}] \\ &= \sum_{k=0}^{\infty} k\mathbf{P}\{\text{an individual of family size } i \text{ has } k \text{ } K^*\text{-type offspring}\} \\ &= \sum_{k=K}^{\infty} k\frac{w_i^k}{k!}e^{-w_i} = \sum_{k=K}^{\infty} \frac{w_i^k}{(k-1)!}e^{-w_i} \\ &= \sum_{j=1}^{\infty} \frac{w_i^j}{(j-1)!}e^{-w_i} - \sum_{j=1}^{K-1} \frac{w_i^j}{(j-1)!}e^{-w_i} \\ &= w_i - \sum_{j=1}^{K-1} m_{ij}. \end{aligned}$$

The entries in the variance matrix $S$ are calculated similarly. For $1 \leq i, j < K$

we obtain

$$\begin{aligned}\sigma_{ij} &= \operatorname{var}{}^{e_i} Z_j(1) \\ &= \sum_{k=0}^{\infty} k^2 \mathbf{P}\{\text{an individual of family size } i \text{ has } k \text{ } j\text{-type offspring}\} - m_{ij}^2 \\ &= j^2 \frac{w_i^j}{j!} e^{-w_i} - \left( j \frac{w_i^j}{j!} e^{-w_i} \right)^2 .\end{aligned}$$

For $j < K$ we have

$$\sigma_{K^* j} = j^2 \frac{w_K^j}{j!} e^{-w_K} - \left( j \frac{w_K^j}{j!} e^{-w_K} \right)^2 ,$$

and finally, for $1 \le i \le K$,

$$\begin{aligned}\sigma_{iK^*} &= \sum_{j=K}^{\infty} \left( j^2 \frac{w_i^j}{j!} e^{-w_i} - m_{ij}^2 \right) \\ &= \sum_{j=K}^{\infty} \sigma_{ij}.\end{aligned}$$

We pointed out that the process $Z(t)$ is a Markov chain. The following classic line evaluates the conditional expectation of the population size vector $Z(t)$ in generation $t$ given the population size vector in the previous generation.

$$\begin{aligned}\mathrm{E}\,[Z_j(t)|Z(t-1)] = \mathrm{E}\,[\sum_{i=1}^{K^*} Z_i(t-1)\eta_{ij}|Z(t-1)] &= \sum_{i=1}^{K^*} \mathrm{E}\,[Z_i(t-1)\eta_{ij}|Z(t-1)] \\ = \sum_{i=1}^{K^*} Z_i(t-1) m_{ij} &= (Z(t-1)M)_j,\end{aligned}$$

because $\eta_{ij}$, the number of $j$-type offspring of an $i$-type individual, is independent of $Z(t-1)$, and $\mathrm{E}\,\eta_{ij} = \mathrm{E}^{\,e_i} Z_j(1)$. Taking expectation and using Mathematical

Induction, we obtain that for all $t \in \mathbf{N}$,

$$\mathrm{E}\,[Z(t)]^T = \mathrm{E}\,[Z(0)]^T M^t,$$

where superscript T denotes matrix transposition. Exploiting the individuals' independent reproduction, we similarly obtain that

$$\mathrm{var}\,[Z(t)|Z(t-1)]^T = Z(t-1)^T S.$$

Hence the variance of the population size is

$$\begin{aligned}
\mathrm{var}\,[Z(t)] &= \mathrm{E}\,[\mathrm{var}\,[Z(t)|Z(t-1)]] + \mathrm{var}\,[\mathrm{E}\,[Z(t)|Z(t-1)]] \\
&= \mathrm{E}\,[Z(t-1)]S + \mathrm{var}\,[Z(t-1)M] \\
&= \mathrm{E}\,[Z(0)]M^{t-1}S + \mathrm{var}\,[Z(t-1)]M^2 \\
&= \mathrm{E}\,[Z(0)](M^{t-1} + M^t + ... + M^{2t-2})S + \mathrm{var}\,[Z(0)]M^{2t}; \text{ if } Z(0) = z_0 \text{ a. s.} \\
&= \mathrm{E}\,[Z(0)](M^{t-1} + M^t + ... + M^{2t-2})S.
\end{aligned}$$

We can transform powers of a matrix into powers of one of its eigenvalues. Let $\lambda$ denote an eigenvalue of the matrix $M$ and $z_0$ be a corresponding eigenvector. Then starting out from a population whose family size structure is $z_0$, that is, $Z(0) = \alpha z_0^T$ almost surely ($\alpha \in \mathbf{R}^+$), we obtain

$$\mathrm{E}\,[Z(t)]^T = \mathrm{E}\,[Z(0)]^T M^t = \alpha z_0^T \rho^t.$$

The Perron-Frobenius Theorem provides an asymptotic-time characterisation even when the initial population vector is not an eigenvector. The mean matrix $M$ is positive and irreducible as $m_{ij} > 0$ for all $i$ and $j$ such that $1 \leq i, j \leq K$. The

powers of such matrices are characterised by their largest eigenvalue in modulus, often referred to as the principal eigenvalue of the matrix. Now, by the Perron-Frobenius Theorem, the multiplicity of the largest eigenvalue, $\rho$ of $M$ is one, and hence there are a unique left eigenvector $u$ and a unique right eigenvector $v$ such that

$$u^T M = \rho u^T,\ u^T v = 1 \text{ and } \sum_{j=1}^{K^*} u_j = 1.$$

moreover, writing $A = (a_{ij})$ for the $K \times K$ matrix defined by $a_{ij} = v_i u_j$, we have

$$M^t = \rho^t A + o(\rho^t).$$

Hence the asymptotic mean population size is given by

$$\mathrm{E}\left[Z(t)^T\right] = \rho^t E[Z(0)^T]A + o(\rho^t).$$

If $\rho > 1$, the long term variance, in turn, can be found:

$$\begin{aligned}\mathrm{var}\left[Z(t)\right]^T &= \mathrm{E}\left[Z(0)\right]^T(\rho^{t-1} + \rho^t + ... + \rho^{2t-2})AS + o(\rho^t) \\ &= \mathrm{E}\left[Z(0)\right]^T \rho^{t-1}\frac{\rho^t - 1}{\rho - 1}\, AS + o(\rho^t).\end{aligned}$$

Likewise, if $\rho = 1$, then

$$\mathrm{var}\left[Z(t)\right]^T = t\,\mathrm{E}\left[Z(0)\right]^T AS,$$

and lastly, if $\rho < 1$, then

$$\operatorname{var}[Z(t)]^T = \mathrm{E}[Z(0)]^T \rho^{t-1} \frac{\rho^t - 1}{\rho - 1} AS + o(\rho).$$

Thus, similarly to the simple single-type Galton-Watson process, all the family size classes increase exponentially if $\rho > 1$ (supercritical case), decrease exponentially if $\rho < 1$ (subcritical case), or remain stable if $\rho = 1$ (critical case).

Although its mean behaviour is not a complete description of the process, these precursory calculations tell about the suitability of our model for the biological problem at hand. If the largest eigenvalue of $M$ is greater than one, then we see an exponential growth in the expected size of the developing population, while, if it is one, then the mean population size remains stable (but the population will die out eventually). Human populations do grow geometrically but with a small common ratio (Malthus, 1798), well approximated by polynomial growth. This is achieved by choosing the weights $w^{(N)}$ such that $\rho = 1 + \frac{\beta}{N}$ for large $N \in \mathbf{N}^+$ and some $\beta(w^{(N)}) \in \mathbf{R}$.

Once the largest eigenvalue of the mean matrix is known, we can decide on the criticality of the branching process and readily describe the first two moments of the process. For large matrices the principal eigenvalue is not easily found, however. In order to keep the calculations simple we specify the reproductive weights $(w_j)$.

**Definition 5.7** (Near-critical weights)**.** *For some given sequence of real numbers $\epsilon$ and natural number $N$, define the weight sequence $w^{(N,\epsilon)}$ by writing $w_i = 1 + \frac{\epsilon_i}{N}$. We will refer to weights of this form as near-critical weights.*

Our choice of weights on the one hand simplifies the calculation of the principal eigenvalue of the mean matrix, and on the other hand it makes possible a diffusion approximation for the rescaled process $N^{-1}Z(\lfloor Nt \rfloor)$ for large values of the

parameter $N$.

We illustrate the above concepts by setting the number of distinct types $K$ to two. In doing so we distinguish between only two types of individuals: those who have siblings and those who grow up alone. In this setting, it is straightforward to calculate the matrix $M$, its principal eigenvalue $\rho$ and the corresponding left and right eigenvalues $u$ and $v$. As we shall see, this mini-model is rich enough: the branching process can be subcritical, critical or supercritical even if we require that $w_1 < 0 < w_2$, in other words, that singletons be handicapped.

**Example 5.8** (Two-type CTF branching process with near-critical weights)**.** In order to avoid cluttering our calculations with a lot of terms of the form $we^{-w}$, we introduce the shorthand notation: $a = w_1 e^{-w_1}$ and $b = w_2 e^{-w_2}$. Then the mean matrix $M_2$ is of the form

$$M_2 = \begin{pmatrix} a & w_1 - a \\ b & w_2 - b \end{pmatrix},$$

with $a = w_1 e^{-w_1} = \left(1 + \frac{\epsilon_1}{N}\right) e^{-1} \left(1 - \frac{\epsilon_1}{N} + O(\frac{1}{N^2})\right) = e^{-1} + O(1/N^2)$ and $b = e^{-1} + O(1/N^2)$ is obtained similarly.

The principal eigenvalue of $M_2$ is the largest root of the characteristic polynomial $c_{M_2}$ associated with the mean matrix.

$$\begin{aligned} c_{M_2}(\lambda) &= \lambda^2 - \operatorname{tr} M_2 \lambda + \det M_2 \\ &= \lambda^2 + (b - a - w_2)\lambda + a w_2 - b w_1 \\ &= \lambda^2 - \left(1 + \frac{\epsilon_2}{N}\right) \lambda + \frac{\epsilon_2 - \epsilon_1}{N} e^{-1} + O(N^{-2}). \end{aligned}$$

Assuming that $\lambda$ is an irrational number, we can take $\lambda$'s unique $N$-ary expansion: there exist $\mu$ and $\kappa \in \mathbf{R}$ such that

$$\begin{aligned}\lambda &= \mu + \frac{\kappa}{N} + O(N^{-2}) \text{ and consequently} \\ \lambda^2 &= \mu^2 + \frac{2\kappa\mu}{N} + O(N^{-2}).\end{aligned}$$

Substituting these expressions for $\lambda$ and $\lambda^2$ in the quadratic equation $c_{M_2}(\lambda) = 0$ and equating the '$N$-ary places' we obtain a simple system of quadratic equations for $\mu$ and $\kappa$.

$$\begin{aligned}\mu^2 + \frac{2\kappa\mu}{N} - \mu - \frac{\epsilon_2}{N}\mu - \frac{\kappa}{N} + \frac{\epsilon_2 - \epsilon_1}{N}e^{-1} + O(N^{-2}) = 0 \\ \mu^2 - \mu + \frac{1}{N}[2\kappa\mu - \epsilon_2\mu - \kappa + (\epsilon_2 - \epsilon_1)e^{-1}] + O(N^{-2}) = 0.\end{aligned}$$

Whence we obtain that

$$\begin{aligned}&\mu^2 - \mu = 0; \qquad \mu(\mu - 1) = 0; \qquad \mu_{1,2} = 0, 1 \text{ and} \\ &2\kappa\mu - \epsilon_2\mu - \kappa + (\epsilon_2 - \epsilon_1)e^{-1} = 0; \qquad \kappa = \frac{(\epsilon_1 - \epsilon_2)e^{-1} + \epsilon_2\mu}{2\mu - 1}; \\ &\kappa_{1,2} = (\epsilon_2 - \epsilon_1)e^{-1}, (\epsilon_1 - \epsilon_2)e^{-1} + \epsilon_2.\end{aligned}$$

Thus the eigenvalues of the mean matrix are $\lambda_1 = \frac{(\epsilon_2 - \epsilon_1)e^{-1}}{N} + O(1/N^2)$ and $\lambda_2 = 1 + \frac{\epsilon_2(1-e^{-1}) + \epsilon_1 e^{-1}}{N} + O(1/N^2)$. The principal eigenvalue is the largest of these two eigenvalues: $\rho = \lambda_2$. The branching process is supercritical, (critical, subcritical) if and only if $\rho > 1$, $(= 1, < 1)$. Because $1 - e^{-1} > 0$, we have just shown that the process is supercritical if $\epsilon_2 > -\epsilon_1/(e-1)$. This demonstrates that our mini example is sufficiently rich. Even if, in this set-up, singleton individuals are disadvantaged, $w_1 < 1 < w_2$, that is, $\epsilon_1 < 0 < \epsilon_2$, the process can still be supercritical, critical or

subcritical depending on the absolute values of the parameters $\epsilon_1$ and $\epsilon_2$.

A similar calculation (matching coefficients in the unique $N$-ary expansions) would yield the corresponding normalised left and right eigenvectors. However this technique becomes exceedingly inconvenient. Instead, for $|\alpha|$ and $|\beta| < N$ we view the expansion $\alpha + \frac{\beta}{N} + O(1/N^2)$ as an element of the field $F := \{\alpha + \frac{\beta}{N} : \alpha, \beta \in \mathbf{R}\}$. By considering its geometric expansion, it is easy to convince oneself that the inverse of the element $\alpha + \frac{\beta}{N}$ is $\frac{1}{\alpha}\left(1 + \frac{1}{N}\frac{\beta}{\alpha}\right)$ in $F$. In this field $N$ is not a number any more. It is merely a symbol. Thus our task simplifies to finding the eigenvalues of a matrix with underlying field $F$: we need not keep track of the error terms during this calculation. In order to show the power of working in a vector space with this underlying field, we find a left eigenvector corresponding to $\rho$, as an example. We are looking for $u_1$ and $u_2 \in F$ such that

$$\begin{pmatrix} u_1 & u_2 \end{pmatrix} \begin{pmatrix} e^{-1} & 1 - e^{-1} + \frac{\epsilon_1}{N} \\ e^{-1} & 1 - e^{-1} + \frac{\epsilon_2}{N} \end{pmatrix} = \rho \begin{pmatrix} u_1 & u_2 \end{pmatrix}.$$

Therefore we obtain

$$\begin{aligned} &e^{-1}(u_1 + u_2) = \rho u_1; \\ &u_1 + u_2 = e\rho u_1; \\ &u_2 = (e\rho - 1)u_1 = \left(e - 1 + \frac{\epsilon_1 + (e-1)\epsilon_2}{N}\right) u_1. \end{aligned}$$

So a left eigenvector belonging to eigenvalue $\rho$ of the matrix $M_2$ is

$$u' = \begin{pmatrix} 1 & e - 1 + \frac{\epsilon_1 + (e-1)\epsilon_2}{N} \end{pmatrix}.$$

An equally simple calculation shows that a corresponding right eigenvector is

$$v' = \begin{pmatrix} 1 & 1 + \frac{\epsilon_2 - \epsilon_1}{N} \end{pmatrix}^T.$$

We may use eigenvectors $u'$ and $v'$ to find the normalised eigenvectors to be

$$u = \begin{pmatrix} 1 - \frac{(1-e^{-1})(\epsilon_2-\epsilon_1)}{N} & 1 + \frac{e^{-1}(\epsilon_2-\epsilon_1)}{N} \end{pmatrix} \text{ and}$$
$$v = e^{-1} \begin{pmatrix} 1 - \frac{e^{-1}\epsilon_1 + (1-e^{-1})\epsilon_2}{N} & e - 1 + \frac{e^{-1}\epsilon_1 + (1-e^{-1})\epsilon_2}{N} \end{pmatrix}^T,$$

subject to our desired set of conditions: $u^T v = 1$ and $u\mathbf{1} = 1$. Four more multiplications in the field $F$ yield the matrix

$$A = vu = e^{-1} \begin{pmatrix} 1 - \frac{2(1-e^{-1})\epsilon_2 + (2e^{-1}-1)\epsilon_1}{N} & e^{-1} - 1 + \frac{\epsilon_1}{N} \\ 1 + \frac{(2e^{-1}-1)\epsilon_2 - 2e^{-1}\epsilon_1}{N} & e^{-1}\left(1 - e + \frac{(2-e^{-1})\epsilon_1 + (e-2+e^{-1})\epsilon_2}{N}\right) \end{pmatrix}.$$

Now the asymptotic results in the previous section have become available to us. For instance, by generation $t$ the mean size of the population has grown by

$$\rho^t = \left(1 + \frac{\epsilon_1 e^{-1} + \epsilon_2(1-e^{-1})}{N} + O(N^{-2})\right)^t = 1 + \frac{t}{N}(\epsilon_1 e^{-1} + \epsilon_2(1-e^{-1})) + O(N^{-2})$$

times of its original size. Were we to measure time in units of the parameter $N$ the growth factor of the population, $\mathrm{E}\left[Z(Nt)^T - Z(0)^T\right]$, is $t(\epsilon_1 e^{-1} + \epsilon_2(1-e^{-1})) + O(1/N)$.

In §1.3 we saw that individuals who have very many siblings tend to have few children themselves, perhaps because they become conscious of the economic/social disadvantage of having been brought up in a large family. We should like our model to cover this situation as well. The least number of different types to capture this

important feature is three: both singletons and individuals of family size at least three are disadvantaged relative to their counterparts with one sibling. Now the principal eigenvalue $\rho$ of the mean matrix $M$ is a real solution to a cubic equation.

The more dimensions we add, the more cumbersome these calculations become. (We should also note that if we assigned unique weights to the first five family sizes then the above recipe would not work. The characteristic polynomial associated with the mean matrix of the process would be of order five, for whose roots there is no analytical formula.) In order to get around this problem, in the next section, we use techniques from functional analysis to find the principal eigenvalue of the mean matrix of higher dimensional branching CTF processes.

## 5.2 The mean operator: the infinite-dimensional case

As the preceding section demonstrates, the truncation of weights stands in our way. It is difficult to conjecture the principal eigenvalue and the corresponding eigenvectors for large values of the cut-off parameter $K$. In this section we show that it is simpler to work with an infinite sequence of near-critical weights. We begin by fixing a sequence $\epsilon \in l^\infty$, and choosing $N \in \mathbf{N}^+$ such that $N \gg \|\epsilon\|_\infty$. The significance of selecting $\epsilon$ to be bounded will be made clear below. As before, we set the weights $w = (w_i)$ to be

$$w_i = 1 + \frac{\epsilon_i}{N}.$$

This time we treat countably many types (family size classes): the infinite-dimensional version of the mean matrix $M$ is the mean operator.

**Definition 5.9** (Mean operator $T$)**.** *The mean operator $T : l^1 \to l^1$ associated with*

*the branching process $Z$ is defined for all $x \in l^1$ by*

$$Tx = \mathrm{E}\,[Z_1 | Z_0 = x].$$

Sometimes we emphasise that $T$ depends on $\epsilon$ by writing $T^\epsilon$. We also observe that, as reproduction in subsequent generations happens independently, we have

$$T^n x = \mathrm{E}\,[Z_n | Z_0 = x].$$

Recalling our notation introduced in Definition 5.2, a single $i$-type individual produces $j$ offspring of type $j$ with probability $p_{ij} = p_j(w_i)$. Relying on the twice differentiability of $p_j$, we use Taylor's theorem to approximate this probability.

$$\begin{aligned} p_{ij} &= p_j\left(1 + \frac{\epsilon_i}{N}\right) = p_j(1) + p_j'(1)\frac{\epsilon_i}{N} + \frac{p_j''(\eta_i)}{2}\frac{\epsilon_i^2}{N^2} \\ &= p_j(1) + p_j'(1)\frac{\epsilon_i}{N} + O_j(N^{-2}), \end{aligned}$$

for some $\eta_i \in \left(1, 1 + \frac{\epsilon_i}{N}\right)$. As in Part I, here $O_j(1/N^2)$ denotes a summable quantity such that $\sum_{j=1}^{\infty} O_j(1/N^2) = O(1/N^2)$. We observe that the term $O_j(1/N^2)$ in the expression for $p_{ij}$ also depends on $i$. But as $\sup|\epsilon_i| < \infty$, $O_j(1/N^2) = O(1/N^2)$ for all $j \in \mathbf{N}^+$.

Thus the mean number of $j$-type offspring of an $i$-type individual, $q_{ij}$, is given by

$$q_{ij} = jp_{ij} = jp_j(1) + jp_j'(1)\frac{\epsilon_i}{N} + O_j(N^{-2}),$$

because an $i$-type individual can only have $j$ many $j$-type children and the probability of this event is $p_{ij}$. The error term is summable over the variable $j$ since the distribution $p = (p_j)$ has finite expectation. Hence the mean operator $T$ is

approximated by

$$(Tx)_j = (\mathrm{E}\,[Z_1|Z_0 = x])_j = \sum_{i=1}^{\infty} q_{ij}x_i$$
$$= jp_j(1)\sum_{i=1}^{\infty} x_i + j\frac{p_j'(1)}{N}\sum_{i=1}^{\infty}\epsilon_i x_i + O(N^{-2}).$$

We use this estimate repeatedly below. The finite-dimensional model studied in §5.1 is a special case of this: let the offspring distribution be Poisson and the weight sequence $w = (1 + \epsilon_i/N)$ with $\epsilon_i = \epsilon_K$ for all $i \geq K$.

**Example 5.10.** If the offspring distribution is Poisson, $p_{ij} = \frac{w_i^j}{j!}e^{-w_i}$, then the mean number of $j$-type offspring of an $i$-type individual is

$$q_{ij} = j\frac{w_i^j}{j!}e^{-w_i} = \frac{j}{j!}\left(1 + \frac{(j-1)\epsilon_i}{N}\right)e^{-1} + O(N^{-2}).$$

Our objective is to find the 'principal eigenvalue' of $T$ and the corresponding eigenspace, as we did for the mean matrix $M$ in the truncated case with $K = 2$ in §5.1. However, as the operator $T$ is infinite-dimensional, it is not clear whether the spectral radius of $T$ is an eigenvalue as it would be the case if $T$ were finite dimensional. Hence we must begin by showing that $T$ is well-behaved.

**Lemma 5.11.** *The mean operator $T : l^1 \to l^1$ is a bounded linear operator provided that $\epsilon \in l^\infty$.*

*Proof.* Because $T$ is an integral operator, it is linear. Let $x \in l^1$ and observe that

$$\begin{aligned}
\|Tx\|_1 &= \sum_{j=1}^{\infty} |\sum_{i=1}^{\infty} q_{ij} x_i|; \text{ because } q_{ij} \geq 0 \\
&\leq \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} q_{ij} |x_i| = \sum_{i=1}^{\infty} |x_i| \sum_{j=1}^{\infty} j p_{ij}; \text{ as } p_j(w) \text{ is parametrised by its mean} \\
&= \sum_{i=1}^{\infty} w_i |x_i|; \text{ as } w_i = 1 + \frac{\epsilon_i}{N} \\
&= \|x\|_1 + \frac{1}{N} \sum_{i=1}^{\infty} \epsilon_i |x_i| \\
&\leq \left(1 + \frac{\sup_i \epsilon_i}{N}\right) \|x\|_1.
\end{aligned}$$

Thus $T$ is a bounded linear operator on $l^1$; and we have $\|T\| \leq 1 + \frac{\|\epsilon\|_\infty}{N}$. □

As we have already discussed above, one normally cannot speak of the greatest (in modulus) eigenvalue of an infinite-dimensional operator. Compactness is exactly the notion we need.

**Definition 5.12** (Compact operator)**.** *We say that an operator $S : X \to X$ is compact on a Banach space $X$ if the image under $S$ of any bounded subset of $X$ is relatively compact in $X$, in other words, its closure is compact in the space $X$.*

Such an operator is automatically bounded, and hence its spectrum is well-defined. The spectrum of a compact operator is countable, and the only possible point of accumulation in it is 0. The set $\sigma(T) \setminus \{0\}$ consists solely of eigenvalues. So if $T$ were compact, its "principal eigenvalue" would be its spectral radius $r_\sigma(T) := \sup_{\lambda \in \sigma(T)} |\lambda|$. (It is unfortunate that the letter $\sigma$ is already in use for the standard deviation of $p(1)$; but it should be clear what $\sigma$ denotes from the context.) Because compact operators are bounded, as we have pointed out above, we now understand

why we insisted that the sequence $\epsilon$ be bounded (and hence $T$ be bounded via Lemma 5.11).

Our mean operator is an integral operator, in one sense it is an 'infinite dimensional matrix'. One might perhaps think that all bounded integral operators are compact. This is far from the truth. Here is a deterrent, proverbial example.

**Example 5.13.** Take a power series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ with a non-zero radius of convergence. Then the 'infinite matrix'

$$A = \begin{pmatrix} a_0 & 0 & 0 & 0 & \cdots \\ a_1 & a_0 & 0 & 0 & \cdots \\ a_2 & a_1 & a_0 & 0 & \cdots \\ a_3 & a_2 & a_1 & a_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

has eigenvectors $v(z) = (1, z, z^2, \dots)$: as $v(z)A = f(z)v(z)$. So, for all $z \in \mathbf{C}$, $f(z)$ is an eigenvalue of $A$. If we now choose $f$ to be the exponential function, which is entire, we see that the spectrum of $A$ is the whole complex plane except one point, the origin. $A$ cannot be compact, as $\sigma(A)$ is not closed, (nor it is countable).

Luckily, $T$ is compact. We show this by applying the following standard theorem.

**Theorem 5.14** (Kreyszig (1989))**.** *Let $X$ be a Banach space and $(S_n)$ a sequence of compact linear operators on the space $X$. If $S_n$ converges to $S : X \to X$ in $B(X)$ as $n \to \infty$, then $S$ is also compact.*

**Corollary 5.15.** *The mean operator $T^{\epsilon}$ is compact, if $\epsilon \in l^{\infty}$.*

*Proof.* Define a sequence of operators $S_n : l^1 \to l^1$ by, for all $x \in l^1$,

$$(S_n x)_j = \begin{cases} \sum_{i=1}^{\infty} q_{ij} x_i & \text{for all } j < n \\ 0 & \text{otherwise.} \end{cases}$$

So $(S_n x)_j = \sum_{i=1}^{\infty} q_{ij}^{(n)} x_i$, where $q_{ij}^{(n)} = q_{ij}$ if $j < n$, 0 otherwise.

Now the set $\{e_1, \ldots e_{n-1}\}$, where $e_k = \delta_{\cdot k}$, is a basis for the linear space $S_n l^1$, the image of the space of absolutely summable sequences under the operator $S_n$. Because $S_n$ is a linear bounded operator with a finite-dimensional range, $S_n$ is compact. In order to see that $T$ is compact, it is sufficient to show that $S_n \to T$ in the space of bounded linear operators on $l^1$, $B(l^1)$, as $n \to \infty$. Let $x \in l^1$ and observe that

$$\begin{aligned} \|(T - S_n)x\|_1 &= \sum_{j=1}^{\infty} |\sum_{i=1}^{\infty} q_{ij} x_i - \sum_{i=1}^{\infty} q_{ij}^{(n)} x_i| \\ &= \sum_{j=1}^{n-1} |\sum_{i=1}^{\infty} q_{ij} x_i - \sum_{i=1}^{\infty} q_{ij} x_i| + \sum_{j=n}^{\infty} |\sum_{i=1}^{\infty} q_{ij} x_i - 0| \\ &\leq \sum_{j=n}^{\infty} \sum_{i=1}^{\infty} q_{ij} |x_i|; \end{aligned}$$

as the series is absolutely convergent, we may rearrange it

$$\begin{aligned} &= \sum_{i=1}^{\infty} |x_i| \sum_{j=n}^{\infty} q_{ij}; \text{ by definition of } q \\ &\leq \left( \sum_{j=n}^{\infty} j p_j(1) + \frac{\|\epsilon\|_\infty}{N} \sum_{j=n}^{\infty} j p_j'(1) + \sum_{j=n}^{\infty} O_j(N^{-2}) \right) \|x\|_1 \\ &=: f(n, N) \|x\|_1. \end{aligned}$$

Taking the supremum over all $x \in l^1$ of norm 1, we obtain that, for all $n \geq 1$,

$$\|T - S_n\| \leq f(n, N).$$

Since $f(n, N)$ is the tail of a convergent series, $f(n, N) \to 0$ as $n \to \infty$ for all fixed $N \in \mathbf{N}^+$. The sequence $(S_n)$ therefore converges to the mean operator in $B(l^1)$. This completes the proof. □

To summarise, the largest eigenvector in modulus of our mean operator $T$ is its spectral radius $\rho$. We now find $\rho$ by first estimating it from below, exhibiting an eigenvector of $T$: then $\rho$ is necessarily not less than the corresponding eigenvalue. Secondly, we use Gelfand's Theorem to show that this lower bound is also an upper bound for $\rho$.

In order to perform the calculations advertised in §5.1 and to find the law of the limit process $\lim_{N\to\infty} N^{-1} Z(\lfloor Nt \rfloor)$, we also need the left eigenvectors corresponding to $\rho$. These, in turn, are the 'right' eigenvectors of $T$'s adjoint operator.

**Proposition 5.16** (The adjoint operator $T^\times$)**.** *The adjoint of the mean operator, $T^\times : l^\infty \to l^\infty$, is given by, for all $y \in l^\infty \cong l^{1\prime}$,*

$$\begin{aligned}(T^\times y)_j &= \sum_{i=1}^{\infty} q_{ji}\, y_i = \sum_{i=1}^{\infty} i\, p_{ji}\, y_i \\ &= \sum_{i=1}^{\infty} i\, p_i \left(1 + \frac{\epsilon_j}{N}\right) y_i.\end{aligned}$$

As $T$ is compact, $T^\times$ is also compact. Moreover, because $\|T^n\| = \|T^{\times n}\|$, applying Gelfand's formula, we deduce that their spectral radii are equal:

$$r_\sigma(T) = \limsup_{n\to\infty} \sqrt[n]{\|T^n\|} = \limsup_{n\to\infty} \sqrt[n]{\|T^{\times n}\|} = r_\sigma(T^\times).$$

Above we used the following result by I. Gelfand (1941).

**Theorem 5.17** (Gelfand's formula)**.** *If $T$ is a bounded linear operator on a Banach*

*spaces, then the spectral radius $r_\sigma(T)$ of $T$ we have*

$$r_\sigma(T) = \lim_{n\to\infty} \sqrt[n]{\|T^n\|}.$$

**Proposition 5.18.** *For some sequences $\gamma$ and $\delta \in l^\infty$ to be determined,*

$$v^{(N)} := \left( jp_j(1) + \frac{p_j(1)\gamma_j}{N} + O_j(N^{-2}) \right)_{j\geq 1} \in l^1$$

*is an eigenvector of $T$ corresponding to the eigenvalue $\rho = 1 + \frac{1}{N}\sum_{j=1}^{\infty} jp_j(1)\epsilon_j + O(N^{-2})$; and*

$$u^{(N)} := \left( 1 + \frac{\delta_j}{N} + O(N^{-2}) \right)_{j\geq 1} \in l^\infty$$

*is an eigenvector of $T^\times$ corresponding to the eigenvalue $\rho$.*

Again, the significance of this result is that hence $\rho$ is a lower bound for $r_\sigma(T)$, but we will see below applying Gelfand's formula that $\rho$ is also an upper bound.

*Proof.* First note that the sequence $v = (jp_j(1))_j$ is an eigenvector of $T^0$ with eigenvalue 1: for all $j \geq 1$,

$$(T^0 v)_j = \sum_{i=1}^{\infty} q^0_{ij} v_i = jp_j(1) \sum_{i=1}^{\infty} ip_i(1) = jp_j(1) \cdot 1 = v_j.$$

We fix a bounded sequence, $\gamma$. We first assume that $v^{(N)} := v + \left(\frac{\gamma_i p_i(1)}{N}\right)_i$ is an eigenvector of the mean operator $T^\epsilon$ and calculate its corresponding eigenvalue whose

$N$-ary expansion is $\alpha + \beta/N + O(1/N^2)$.

$$\begin{aligned}(T^\epsilon v^{(N)})_i &= \sum_{i=1}^{\infty} q_{ij}^{\epsilon} v_i^{(N)} = \sum_{i=1}^{\infty} jp_j \left(1 + \frac{\epsilon_i}{N}\right) v_i^{(N)} \\ &= \sum_{i=1}^{\infty} \left(jp_j(1) + jp_j'(1)\frac{\epsilon_i}{N} + O_j(N^{-2})\right)\left(ip_i(1) + \frac{\gamma_i p_i(1)}{N}\right) \\ &= jp_j(1)\sum_{i=1}^{\infty} ip_i(1) + \frac{jp_j'(1)}{N}\sum_{i=1}^{\infty} i\epsilon_i p_i(1) + \frac{jp_j(1)}{N}\sum_{i=1}^{\infty}\gamma_i p_i(1) + O(N^{-2}) \\ &= \left(1 + \frac{\beta}{N}\right)\left(jp_j(1) + \frac{\gamma_j}{N}p_j(1)\right) = jp_j(1) + \frac{j\beta p_j(1)}{N} + \frac{\gamma_j p_j(1)}{N} + O(N^{-2}),\end{aligned} \tag{5.1}$$

where the penultimate equality holds by assumption of the eigenvalue property. Cancellation on the two sides of that equality gives

$$(j\beta + \gamma_j)p_j(1) = jp_j'(1)\sum_{i=1}^{\infty} i\epsilon_i p_i(1) + jp_j(1)\sum_{i=1}^{\infty}\gamma_i p_i(1). \tag{5.2}$$

Summing both sides from 1 to $n$ now yields

$$\beta\sum_{j=1}^{n} jp_j(1) + \sum_{j=1}^{n}\gamma_j p_j(1) = \sum_{j=1}^{n} jp_j'(1)\cdot\sum_{i=1}^{\infty} i\epsilon_i p_i(1) + \sum_{j=1}^{n} jp_j(1)\cdot\sum_{i=1}^{\infty}\gamma_i p_i(1).$$

The above series converge. Taking the limits as $n \to \infty$, we finally obtain that

$$\beta\cdot 1 + \sum_{j=1}^{\infty}\gamma_j p_j(1) = 1\cdot\sum_{i=1}^{\infty} i\epsilon_i p_i(1) + 1\cdot\sum_{i=1}^{\infty}\gamma_i p_i(1).$$

Rearranging the equation, we find that

$$\beta = \sum_{i=1}^{\infty} i\epsilon_i p_i(1).$$

The key quantity $\beta$ does not depend on $\gamma$. Thus the above calculation shows that

if we start by defining $\beta$ to be $\sum_{j=1}^{\infty} j\epsilon_j p_j(1)$, and subsequently substitute it into Equation (5.1), we see that $T^{\epsilon}v^{\gamma} = \left(1 + \frac{\beta(\epsilon)}{N}\right) v^{\gamma} + O(N^{-2})$, for some bounded sequence $\gamma$ that satisfies Equation (5.2). In other words, $v^{\gamma}$ is an eigenvector of the mean operator with eigenvalue $\rho := 1 + \beta/N + O(1/N^2)$. It remains to recover the sequence $\gamma$.

Rearranging Equation (5.2), we see that $\gamma$ needs to satisfy

$$p_j(1)\gamma_j = \beta(jp'_j(1) - jp_j(1)) + jp_j(1)\sum_{i=1}^{\infty} p_i(1)\gamma_i. \tag{5.3}$$

Choosing $\gamma_j = j\beta(p'_j(1)/p_j(1) - 1)$, we obtain that

$$\sum_{j=1}^{\infty} p_j(1)\gamma_j = \beta\sum_{j=1}^{\infty} j(p'_j(1) - p_j(1)) = \beta(1-1) = 0,$$

because the distribution $p$ is parametrised by its mean, and hence $\sum_{i=1}^{\infty} jp'_j(w) = w' = 1$. Thus our choice for $\gamma$ satisfies Equation (5.3), and now the calculation in (5.1) shows that the eigenvectors belonging to the eigenvalue

$$\rho = 1 + \frac{\beta}{N} + O(N^{-2}) = 1 + \frac{1}{N}\sum_{i=1}^{\infty} i\epsilon_i p_i(1) + O(N^{-2})$$

are the scalar multiples of the sequence

$$v_j^{(N)} = jp_j(1) + \frac{j}{N}\left((p'_j(1) - p_j(1))\sum_{i=1}^{\infty} i\epsilon_i p_i(1)\right) + O(N^{-2}).$$

The proof of the claim about the adjoint operator is similar. □

We see that $\rho + O(1/N^2)$ lies in the point spectrum of $T$. Thus $T$'s spectral radius is at least $\rho + O(1/N^2)$. Next, using Gelfand's formula, we show that $r_{\sigma}(T)$ is also at most $\rho + O(1/N^2)$.

**Proposition 5.19.** *Let* $\beta(\epsilon) = \sum_{i=1}^{\infty} i\epsilon_i p_i(1)$*, for any bounded sequence* $\epsilon$*. The spectral radius of* $T^{\epsilon}$ *is*

$$r_\sigma(T) = 1 + \frac{\beta}{N} + O(N^{-2}) = \rho + O(N^{-2}).$$

*Proof.* We first use Mathematical Induction to prove that

$$\|T^n x\|_1 \leq \left(1 + \frac{(n-1)\beta}{N}\right)\|x\|_1 + \frac{1}{N}\sum_{k=1}^{\infty}\epsilon_k|x_k| + O(N^{-2}).$$

The base case ($n = 1$) has been established in Lemma 5.11. For the inductive step we observe that

$$\begin{aligned}
\|T^{n+1}x\|_1 &\leq \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} q_{ij}|(T^n x)_i| = \sum_{i=1}^{\infty}|(T^n x)_i|\sum_{j=1}^{\infty} q_{ij};\ \text{as } q_{ij} = jp_j(w_i) \\
&= \sum_{i=1}^{\infty}|(T^n x)_i|\sum_{j=1}^{\infty} jp_j\left(1 + \frac{\epsilon_i}{N}\right);\ \text{as } p(w) \text{ is parametrised by its mean } w \\
&= \sum_{i=1}^{\infty}\left(1 + \frac{\epsilon_i}{N}\right)|(T^n x)_i| \\
&= \|T^n x\|_1 + \frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i|(T^n x)_i|;\ \text{by the induction hypothesis} \\
&\leq \left(1 + \frac{(n-1)\beta}{N}\right)\|x\|_1 + \frac{1}{N}\sum_{k=1}^{\infty}\epsilon_k|x_k| + \frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i|(T^n x)_i| + O(N^{-2}).
\end{aligned}$$

It remains to show that

$$\frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i|(T^n x)_i| \leq \frac{\beta}{N}\|x\|_1 + O(N^{-2}).$$

To verify this we observe that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i|(T^n x)_i| &\leq \frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i\sum_{j=1}^{\infty}q_{ji}|(T^{n-1}x)_j| \\
&= \frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i\sum_{j=1}^{\infty}ip_i\left(1+\frac{\epsilon_j}{N}\right)|(T^{n-1}x)_j| \\
&= \frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i\sum_{j=1}^{\infty}\left(ip_i(1)+ip_i'(1)\frac{\epsilon_j}{N}+O_i(N^{-2})\right)|(T^{n-1}x)_j| \\
&= \frac{1}{N}\sum_{i=1}^{\infty}i\epsilon_i p_i(1)\sum_{j=1}^{\infty}|(T^{n-1}x)_j| + \frac{1}{N}\sum_{i=1}^{\infty}\epsilon_i\sum_{j=1}^{\infty}O_i(N^{-1})|(T^{n-1}x)_j| \\
&= \frac{1}{N}\sum_{i=1}^{\infty}i\epsilon_i p_i(1)\|T^{n-1}x\|_1 + O(N^{-2}); \\
&\qquad \text{by definition of } \beta \text{ and the induction hypothesis} \\
&\leq \frac{\beta}{N}\|x\|_1 + O(N^{-2}).
\end{aligned}$$

This establishes the case, and completes the proof of our claim. The definition of the norm of an operator gives us an upper bound for $\|T^n\|$.

$$\begin{aligned}
\|T^n\| &= \sup_{x\neq 0}\frac{\|T^n x\|_1}{\|x\|_1} \\
&\leq \sup_{x\neq 0}\left\{1+\frac{(n-1)\beta}{N}+O(N^{-2})+\frac{1}{N}\frac{\sum_{k=1}^{\infty}\epsilon_k|x_k|}{\|x\|_1}\right\} \\
&= 1+\frac{(n-1)\beta}{N}+O(N^{-2})+\frac{1}{N}\sup_{x\neq 0}\frac{\sum_{k=1}^{\infty}\epsilon_k|x_k|}{\|x\|_1}.
\end{aligned}$$

Writing $A$ for the supremum in the last line and using Gelfand's formula, finally we

obtain that

$$\begin{aligned} r_\sigma(T) &= \lim_{n\to\infty} \sqrt[n]{\|T^n\|} \text{ Gelfand's formula;} \\ &\leq \lim_{n\to\infty} \left(1 + \frac{(n-1)\beta + A}{N} + O(N^{-2})\right)^{\frac{1}{n}} \text{; Binomial Theorem} \\ &= \lim_{n\to\infty} \left(1 + \frac{n-1}{n}\cdot\frac{\beta}{N} + \frac{1}{n}\cdot\frac{A}{N} + O(N^{-2})\right) \\ &= 1 + \frac{\beta}{N} + O(N^{-2}). \end{aligned}$$

It follows therefore that $r_\sigma(T) = 1 + \beta/N + O(1/N^2) = \rho + O(1/N^2)$ from the previous proposition. □

**Proposition 5.20.** *Let $v = \lim_{N\to\infty} v^{(N)}$ and $u = \lim_{N\to\infty} u^{(N)}$. Then $\|v\|_1 = 1$ and applying the functional $u \in l^{1'} \cong l^\infty$ we obtain $u(v) = \sum_{j=1}^\infty u_j v_j = 1$, too.*

*Proof.* This is a straightforward calculation: $v = (ip_i(1))$ and $u = (1, 1, 1, \dots)$. Whence we obtain that both $\|v\|_1$ and $u(v)$ equal $\sum_{i=1}^\infty ip_i(1)\cdot 1 = 1$ by the property of the offspring distribution laid out in Definition 5.2. □

Using a constant tail sequence $\epsilon = (\epsilon_1, \dots \epsilon_K, \epsilon_K, \dots)$ in our model, we return to a $K$-truncated population process $Z$. Now we are able to calculate the principal eigenvalue of the mean matrix, using the above results; $\epsilon$ and, in turn, our weight sequence $w = (1 + \epsilon_i/N)_i$ are automatically bounded. We demonstrate this strategy with the case of Poisson offspring numbers.

**Example 5.21** (Poisson offspring)**.** We take the offspring probability mass function $p(w_i) = (p_j(w_i))_j$ to be of Poisson variables with mean $w_i$. In other words,

$$p_j(w_i) = \frac{w_i^j}{j!} e^{-w_i},$$

for all $j \in \mathbf{N}$. We then make the weight sequence $w$ have constant tail: $w_i = 1+\epsilon_K/N$ for all $i \geq K$. Now Propositions 5.18 and 5.19 give us that the spectral radius of the mean operator $T^\epsilon$ is $1 + \beta/N + O(N^{-2})$, where

$$\begin{aligned}\beta &= \sum_{i=1}^{\infty} i\epsilon_i p_i(1) = \sum_{i=1}^{\infty} \frac{i}{i!}\epsilon_i e^{-1} = \sum_{i=1}^{K} \frac{i}{i!}\epsilon_i e^{-1} + \epsilon_K \sum_{i=K+1}^{\infty} \frac{i}{i!} e^{-1} \\ &= \sum_{i=1}^{K} \frac{i}{i!}\epsilon_i e^{-1} + \epsilon_K \left( \sum_{i=1}^{\infty} \frac{i}{i!} e^{-1} - \sum_{i=1}^{K} \frac{i}{i!} e^{-1} \right) \\ &= \sum_{i=1}^{K} \frac{i(\epsilon_i - \epsilon_K)}{i!} e^{-1} + \epsilon_K;\end{aligned}$$

and the corresponding limit eigensequence is $v$ given by $v_j = jp_j(1) = \frac{j}{j!}e^{-1}$. Thus after truncation the limit eigenvector of the mean matrix belonging to its principal eigenvalue is $\hat{v}$ whose entries are given by $\hat{v}_j = v_j = \frac{j}{j!}e^{-1}$ for $j \leq K-1$ while $\hat{v}_K = 1 - \sum_{j=1}^{K-1} \frac{j}{j!}e^{-1}$. This way we have reduced the infinite-dimensional branching process to a finite-dimensional one.

Setting the number of different family size distributions, $K$ to two, we can check now our results in Example 5.8. We have shortened a four-page-long calculation there to just a few lines here. Of course, the method above works with any $K \in \mathbf{N}^+$, neither are we limited to the Poisson distribution as the next example shows.

**Example 5.22** (Geometric offspring)**.** Choosing geometric offspring distributions, we first re-parametrise $Geom(\lambda)$ by its mean to calculate $\beta$ and the limit eigenvector $\hat{v}$. The distribution is supported on $\{0, 1, 2, \dots\}$, thus its mean equals $(1-\lambda)/\lambda =: w$; whence $\lambda = 1/(1+w)$. So the re-parametrised probability mass function of the offspring distribution with mean $w$ is given by

$$p_j(w) = \left(1 - \frac{1}{1+w}\right)^j \frac{1}{1+w}$$

for all $j \in \mathbf{N}$. Now we simply obtain that

$$\beta = \sum_{i=1}^{\infty} i\epsilon_i p_i(1) = \sum_{i=1}^{\infty} i\epsilon_i \left(1 - \frac{1}{2}\right)^j \frac{1}{2} = \sum_{i=1}^{\infty} \frac{i\epsilon_i}{2^{i+1}}.$$

The $N$-limit of the eigenvector belonging to the principal eigenvalue of the mean matrix, in turn, is given by its components $\hat{v}_j = \frac{j}{2^{j+1}}$ for $1 \leq j < K$, while $\hat{v}_{K^*} = 1 - \sum_{j=1}^{K-1} \frac{j}{2^{j+1}}$.

## 5.3 Diffusion approximation of family and allele frequencies

So far we have seen that the mean and variance matrices associated with the process can be used in determining the first two moments of the population size. This involves calculating powers of the mean matrix. Using the mean matrix' largest eigenvalue and the corresponding normalised eigenvectors we can describe the long-time asymptotic behaviour of the moments circumventing matrix powers.

In general, however, we are interested not only in the first few moments of a stochastic process but in its whole law. Important quantities, such as time to extinction of the branching process, cannot be determined using only the mean and variance of the population sizes throughout the generations.

In this section, we equip our CTF branching process $Z(t)$ with near-critical weights (Definition 5.7) and study the rescaled population process $X^N(t) = \frac{1}{N} Z([Nt])$. A. Joffe and M. Metivier (1986) showed that the sequence $(X^N(t))_{N \geq 1}$ converges weakly to a diffusion $X(t)$. This also means that finite time statistics of the processes $X^N$ converge to those of the diffusion $X$. Lastly, it is the populations surviving until today in which we are really interested. We connect the thesis to recent development in the theory of branching processes: Budhiraja and Reinhold (2010) established the important property that the asymptotic distributions of $X^N(t)$ conditioned to sur-

vive converges to the distribution of the similarly conditioned diffusion $X(t)$. We make good use of these convergence results, as diffusions are easier to manipulate than branching processes. Finally, at the end of the section, we calculate the dynamics of allele frequencies in our populations.

Budhiraja and Reinhold (2010) gives four sufficient conditions under which the rescaled $K$-type process $X^N(t)$ converges to a vector valued but one-dimensional diffusion.

**Condition 5.23** ($x \log x$)**.** $\mathrm{E}^{\boldsymbol{1}}[\|Z^{(N)}(1)\|_1 \log \|Z^{(N)}(1)\|_1] < \infty$ *for each* $N \in \mathbf{N}^+$. *Here the expectation is taken on the condition that* $Z(0) = (1, 1, \ldots 1)$.

The $x \log x$-condition is equivalent to the non-degeneracy of the limit of the natural martingale associated with a branching process. In our case the offspring distributions have finite variance, thus this condition holds. The second condition concerns the second and third moments of the offspring numbers.

**Condition 5.24.** *The second moments of the offspring numbers are uniformly bounded away from 0 in* $N$*, while their third moments are uniformly bounded in* $N$.

This, of course, holds in the case where the offspring numbers are Poisson distributed ($p_j(w) = w^j/j!\, exp(-w)$) with mean parameter $w_i = 1+\epsilon_i/N$ depending on the family size of the parent, provided that $\|\epsilon\|_\infty < \infty$. But should we use our more general offspring distributions ($p_j(w)$) specified at the beginning of the chapter, we need to make sure that their second and third moments are well-behaved.

**Condition 5.25.** *The mean matrix* $M^{(N)}$ *is of the form* $M^{(N)} = M + C^{(N)}/N$, *with* $M$ *strictly positive and* $\lim_{N\to\infty} C^{(N)} = C$. *The greatest eigenvalues* $\rho_N$ *of the matrices* $C^{(N)}$ *in turn are of the form* $\rho_N = 1 + c_N/N$, *with* $c_N \in (-N, 0)$ *and* $\lim_{N\to\infty} c_N = c \in (-\infty, 0)$. *We also require that* $M$ *has maximal eigenvalue*

*1 with corresponding left eigenvector* $v = \lim_{N\to\infty} v^{(N)}$ *and right eigenvector* $u = \lim_{N\to\infty} u^{(N)}$.

We note that, as a consequence of the this condition, $v^T C u = c$. In the previous section we have found that $\rho_N = 1+\beta/N+O(1/N^2)$ and thus that $c_N = \beta+O(1/N)$. So $c = \beta$. We also obtained there that the limiting left and right eigenvectors are given by $v_j = jp_j(1)$ and $u_j = 1$ for $1 \le j < K$, and by $v_{K^*} = 1 - \sum_{i=1}^{K-1} v_i$ and $u_{K^*} = 1$.

**Definition 5.26** (Covariance matrix of an $l$-type individual's offspring)**.** *Define the* $K \times K$ *covariance matrix* $\sigma(l) = (\sigma_{ij}^{(N)}(l))$ *by writing, for each* $1 \le i, j \le K$,

$$\sigma_{ij}^{(N)}(l) = \sum_{r\in\mathbf{N}^K} (r_i - m_{li}^{(N)})(r_j - m_{lj}^{(N)})p^{(N)}(e_l, r).$$

Here, as before, $p^{(N)}(e_l, r)$ denotes the probability that an $l$-type individual leaves behind the offspring vector $r$, $r_i$ being the number of her $i$-type children.

The variances enter through the following quadratic form.

**Condition 5.27** (Convergence of the covariance matrices)**.** $\lim_{N\to\infty} \sigma^{(N)}(l) = \sigma(l)$ *for all types* $1 \le l \le K$.

In order to state the limit theorem, we also define the positive scalar $Q$ by the quadratic form

$$Q = \frac{1}{2}\sum_{l=1}^{K} v_l u^T \sigma(l) u.$$

**Theorem 5.28** (Feller approximation, Budhiraja and Reinhold (2010))**.** *Assume that Conditions 5.24, 5.25 and 5.27 hold and suppose that the distribution of* $X^{(N)}(0)$ *converges to some* $\mu \in m\mathbf{B}^K$. *Let* $\mu_1 \in m\mathbf{B}$ *be given as*

$$\mu_1(A) = \mu\{x \in \mathbf{R}_+^K : x^T u \in A\}$$

*for all* $A \in \mathbf{B}$; *and let* $\zeta^{(N)} = {X^{(N)}}^T u^{(N)}$, *the total population size.*

*Then* $\zeta^{(N)}$ *converges in distribution to the unique in law diffusion* $\zeta$ *with initial distribution* $\mu_1$ *and generator* $L$ *given by*

$$(Lf)(x) = cxf'(x) + Qxf''(x),$$

*for* $f \in C_K^\infty(\mathbf{R}^+)$ *and* $x \in \mathbf{R}^+$. *Furthermore, for any* $t_0 \in (0, \infty)$, *the process* $X^{(N,0)}$, *defined by* $X^{(N,0)}(t) = X^{(N)}(t_0 + t)$ *for* $t \geq 0$, *converges in distribution to* $X^{(0)} := v\zeta^{(0)}(t) = \zeta(t_0 + t)$, *for* $t \geq 0$.

In the above statement $C_K^\infty(\mathbf{R}^+)$ denotes the infinitely many times differentiable functions with compact support on the positive half line. The sets $\mathbf{B}$ and $m\mathbf{B}$ denote the family of Borel sets and the set of Borel-measurable functions, respectively. The theorem informally states that although the limit process $\lim_{N\to\infty} N^{-1}Z(\lfloor Nt \rfloor)$ is vector valued, it is a one-dimensional Feller diffusion that is supported on the ray $\{\lambda v : \lambda \in \mathbf{R}^+\}$. We have calculated $c = \beta = \sum_{i=1}^\infty i\epsilon_i p_i(1)$; it remains to evaluate $Q$. Thus the theorem applies for our branching process provided Condition 5.24 on the second and third moments holds.

We begin calculating the scalar $Q$ by recalling that an $i$-type individual's offspring vector is $x = je_j$, with probability $p(e_i, x) = p_{ij} = p_j(w_i) = p_j\left(1 + \epsilon_i/N\right)$, where $(e_j)$ denotes the Schauder basis of $l^1$:

$$e_1 = (1, 0, 0, 0, \dots)$$

$$e_2 = (0, 1, 0, 0, \dots)$$

$$e_3 = (0, 0, 1, 0, \dots)$$

etc. First we take limit of the covariance matrix of an $l$-type individuals' offspring.

The off-diagonal entries are of the form

$$\begin{aligned}
\sigma(l)_{ij} &= \lim_{N\to\infty} \sigma_{ij}^{(N)}(l) = \sum_{r\in l^1} (r_i - q_{li})(r_j - q_{lj}) p(e_l, r) \\
&= \sum_{r=ke_k} (r_i - ip_i(1))\,(r_j - jp_j(1))\,p(e_l, r);\ \text{definition of } p(e_l, r) \\
&= \sum_{k=1,k\neq i,j}^{\infty} ip_i(1)jp_j(1)p_k(1) - [(i - ip_i(1))\,jp_j(1)p_i(1)] - \\
&\quad - [(j - jp_j(1))\,ip_1(1)p_j(1)] \\
&= ip_i(1)jp_j(1) - ijp_j(1)p_i(1) - jip_i(1)p_j(1) \\
&= -ip_i(i)jp_j(1).
\end{aligned}$$

In turn, for all family sizes $l \geq 1$, the diagonal entries

$$\begin{aligned}
\sigma_{ii}(l) &= \sum_{k=0,k\neq i}^{\infty} (ip_i(1))^2 p_k(1) + (i - ip_i(1))^2\, p_i(1) \\
&= i^2 p_i(1) - i^2 p_i(1)^2.
\end{aligned}$$

Thus, as one would expect, each individual has the same limiting offspring covariance matrix. We observe that

$$\begin{aligned}
\sum_{j=1}^{\infty} \sigma_{ij} &= \sigma_{ii} + \sum_{j=1,j\neq i}^{\infty} \sigma_{ij} \\
&= i^2 p_i(1) - \sum_{j=1}^{\infty} ip_i(1)jp_j(1) = i(i-1)p_i(1).
\end{aligned}$$

Whence $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sigma_{ij} = \sum_{i=1}^{\infty} i(i-1)p_i(1) =: \sigma^2$, where we introduce the variable $\sigma^2$ to denote the second factorial moment of the mean-one offspring distribution $p(1)$. As the expectation is one, $\sigma^2$ equals the offspring variance. Substituting $\sigma$

into the quadratic form in 5.27, we therefore obtain that

$$\begin{aligned} Q &= \frac{1}{2}\sum_{l=1}^{\infty} v_l u^T \sigma(l) u = \frac{1}{2}\sum_{l=1}^{\infty} l p_l(1) \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} 1 \cdot \sigma_{ij}(l) \cdot 1 \\ &= \frac{\sigma^2}{2}\sum_{l=1}^{\infty} l p_l(1) = \frac{\sigma^2}{2}. \end{aligned}$$

**Definition 5.29** (Yaglom distribution)**.** *Let $(Y_t)$ be a process taking values in a topological space $S$. If $y \in S \setminus \{0\}$, we say that the Yaglom distribution of the process $Y$ started at the point $y$ is the probability measure $\mu$ such that, as $t \to \infty$,*

$$\mathbf{P}^y(Y_t \in \cdot | t < T_Y < \infty) \to \mu(\cdot),$$

*where the stopping time $T_Y$, defined by $T_Y = \inf\{t \in \mathbf{R}^+ : Y_t = 0\}$, is the extinction time of $Y$.*

Now let $\nu$ denote the exponential distribution with density function

$$f(x) = |\beta| Q^{-1} \exp(-|\beta| Q^{-1} x), \quad x \in \mathbf{R}^+.$$

Because $\zeta$ is a Feller diffusion, its Yaglom distribution is $\nu$ by Proposition 2.3.2.1 of Lambert (2008).

The following theorem relates the Yaglom distributions of the rescaled branching processes $X_t^{(N)} = N^{-1} Z(\lfloor Nt \rfloor)$ to the Yaglom distribution of the limit process $v\zeta_t$.

**Theorem 5.30** (Budhiraja and Reinhold (2010))**.** *The Yaglom distributions of the scaled population processes $(X_t^{(N)})$ converge to the Yaglom distribution of $(v\zeta_t)$, as $N \to \infty$, $v$ being the right limit eigenvector of the mean matrix $M^{(N)}$.*

We now understand the sequence of Yaglom distributions of the rescaled branching processes $X^{(N)}$ sufficiently to calculate its limit. This is relevant as human pop-

ulations which may exhibit cultural transmission of fitness all survive today. As a result, we must describe them with a branching process which is conditioned on non-extinction.

**Example 5.31** (Poisson offspring, continuation of Example 5.21)**.** Assuming that the offspring distributions of individuals of family size $i$ is Poisson with parameter $\left(1+\frac{\epsilon_i}{N}\right)$ where $\epsilon_i = \epsilon_K$ for all $i \geq K$, we calculated the value of $\beta$ and $v$ in Example 5.21. In turn, the limit aggregated variance

$$Q = \frac{1}{2}\sum_{i=1}^{\infty} i(i-1)p_i(1) = \frac{1}{2}\sum_{i=1}^{\infty} \frac{i(i-1)}{i!}e^{-1} = \frac{1}{2}.$$

Condition 5.24 on the second and third moments of our Poisson distributed offspring variables is satisfied, thus the limit process $\lim_{N\to\infty} N^{-1}Z^{(\epsilon)}(\lfloor Nt \rfloor) \simeq \zeta_t v$, where $\zeta$ is a Feller diffusion given by the stochastic differential equation

$$d\zeta_t = \sum_{i=1}^{\infty} \frac{\epsilon_i e^{-1}}{(j-1)!}\zeta_t\, dt + \sqrt{\zeta_t}\, dB_t,$$

where $B(t)$ is a standard Brownian motion. Hence the Yaglom density of $\zeta$ is $f : \mathbf{R}^+ \to \mathbf{R}^+$ defined by, for all $x \in \mathbf{R}^+$,

$$f(x) = 2|\sum_{i=1}^{\infty} \frac{\epsilon_i}{(j-1)!}e^{-1}| \exp\left(-2|\sum_{i=1}^{\infty} \frac{\epsilon_i}{(j-1)!}e^{-1}|x\right).$$

Budhiraja and Reinhold's theorem assures us that the Yaglom distribution of $v\zeta$ is also the limit of the Yaglom distributions of the rescaled branching processes $N^{-1}Z^{(\epsilon)}(\lfloor Nt \rfloor)$ as $N \to \infty$.

**Example 5.32** (Geometric offspring, continuation of Example 5.22)**.** If the offspring number of an individual whose family size is $i$ has geometric distribution with mean

$w_i = 1 + \epsilon_i/N$, then, as we calculated in Example 5.22, $p_{ij} = p_j(w_i) = (1 - 1/(1 + w_i))^j/(1 + w_i)$ and $\beta = \sum_{i=1}^{\infty} \frac{i\epsilon_i}{2^{i+1}}$. The $N$-limit of the eigenvector corresponding to the principal eigenvalue of the mean operator is $v = (j/2^{j+1})$. It only remains to determine the limit aggregated variance

$$Q = \frac{1}{2}\sum_{j=1}^{\infty} j(j-1)p_j(1) = \frac{1}{4}\sum_{j=2}^{\infty} \frac{j(j-1)}{2^j} = \frac{1}{4}\,\frac{2x^2}{(1-x)^3}\Big|_{x=1/2} = 1.$$

The third moments are bounded in $N$, so Theorem 5.28 gives the diffusion approximation

$$d\zeta_t = \sum_{i=1}^{\infty} \frac{i\epsilon_i}{2^{i+1}}\zeta_t\,dt + \sqrt{2\zeta_t}\,dB_t$$

for the total rescaled size of the branching process, and $v_j\zeta_t = j\zeta_t/2^{j+1}$ for the $j$th component of the rescaled population vector.

We close the section by applying Theorem 5.28 to describe the change in allele frequencies throughout the generations. Here we are not interested in processes conditioned on non-extinction because while whole populations cannot go extinct, alleles can.

Observe that the total population size of a given allele is $\zeta_t^{(N)}$. Assuming that the population has two competing alleles $A$ and $a$, we write $X_1^{(N)} = \zeta_{t,1}^{(N)}$ for the number of $A$-type individuals and $X_2^{(N)} = \zeta_{t,2}^{(N)}$ for the number of $a$-types, where $\zeta_{t,1}^{(N)}$ and $\zeta_{t,2}^{(N)}$ are two independent realisations of $\zeta_t^{(N)}$. A careful application of Itô's Lemma with the function $g(t, x) = x_1/(x_1 + x_2)$ gives us the allele frequency dynamics. This calculation is classic in the literature.

**Proposition 5.33.** *Let $B_1$ and $B_2$ be two independent standard Brownian motions. The proportion of $A$-type alleles, stopped whenever $A$ fixes or dies out, converges*

*weakly to the diffusion $p_t$ given by*

$$dp_t = \sigma \left( (1-p_t)\sqrt{p_t} dB_1(t) - p_t \sqrt{1-p_t} dB_2(t) \right) / \sqrt{s_t}$$
$$ds_t = \beta s_t dt + \sigma \sqrt{p_t s_t} dB_1(t) + \sigma \sqrt{(1-p_t)s_t} dB_2(t),$$

*where $s^{(N)} = X_1^{(N)} + X_2^{(N)}$ is the total size of the population. Therefore, by Theorem 8.4.3 of Øksendal (2003), $p_t$ is a weak solution to the Wright-Fisher diffusion with a random environment:*

$$dp_t = \sigma \sqrt{\frac{p_t(1-p_t)}{s_t}} dB_t$$
$$ds_t = \beta s_t dt + \sigma \sqrt{s_t} dB_t.$$

**Remark 5.34.** We notice that conditioning the process $p_t$ on the population size $s_t$ being constant 1, one recovers the Wright-Fisher equation $dp_t = \sqrt{p_t(1-p_t)} dB_t$ (provided we choose the offspring distributions to be Poisson, whence $\sigma = 1$.)

## 5.4 An extension with weight sequences of the form $w = (1 + j\alpha/N)$

We recall that in our branching process model for cultural transmission of fitness an individual of family size $i$ has $j$ offspring with probability $p_{ij} = p_j(w_i)$, where the distribution $p_j(w_i)$ is parametrised by its mean:

$$\sum_{j=1}^{\infty} j p_j(w_i) = w_i,$$

where the sequence $w = (w_i)$ is the reproductive weight sequence associated with the process. In §§5.2 and 5.3 we assumed that the weights $w$ are of the form

$$w_i = 1 + \frac{\epsilon_i}{N}$$

for some sequence $\epsilon \in l^\infty$ and a large $N \in \mathbf{N}^+$. Our main result was that the spectral radius of the multi-type branching process' mean operator $T^{(N)}$ is

$$r_\sigma(T^{(N)}) = 1 + \frac{\beta}{N} + O(N^{-2}) = 1 + \frac{1}{N}\sum_{i=1}^{\infty} i\epsilon_i p_i(1) + O(N^{-2})$$

for any bounded sequence $\epsilon$; while, because of the finite covariance condition, we must have had

$$\sum_{i=1}^{\infty} i(i-1)p_i(1) < \infty.$$

These two facts together suggest that we may allow weights of the form

$$w_i = 1 + \frac{\epsilon_i}{N} = 1 + \frac{i\alpha_i}{N}, \qquad \alpha \in l^\infty.$$

In order to adopt these increasing weights, we need a slightly more restrictive condition on the second moment of the offspring number: for each family size $i$, we require that the function

$$\sum_{j=1}^{\infty} j^2 p_j(w_i)$$

is twice differentiable with respect to $w$ on the interval $|w| < \|\alpha\|_\infty$, as a consequence,

$$\sum_{j=1}^{\infty} j^2 p_j'(1) < \infty.$$

In this section we first show that instead of the whole space $l^1$, we can work in a subspace of $l^1$ in which series of the form $\sum_{i=1}^{\infty} i|x_i|$ converge. Secondly, we show that, for increasing weight sequences of the above type, the spectral radius of the restriction of the mean operator to this subspace is still $1+\beta/N+O(1/N^2)$. We close the section by providing an analogue of Theorem 5.28, the Feller approximation of the re-scaled branching process $N^{-1}Z(\lfloor Nt \rfloor)$ for large values of $N$.

Lastly we note that, for large $N$ and small $\alpha$, setting $\alpha_i = \alpha$,

$$w_i = 1 + \frac{i\alpha}{N} \approx \left(1 + \frac{\alpha}{N}\right)^i.$$

The latter quantity is the weight assigned to individuals of family size $i$ (up to a Gamma distributed random factor) in the constant population size model of Sibert, Austerlitz and Heyer (see §2.3 or Sibert et al. (2002)); and exactly our choice for a more simple increasing weight sequence in §4, Part II.

**Definition 5.35** (The space $l^1_*$)**.** *We define the subspace $l^1_*$ of the Banach space of*

*absolutely summable sequences by*

$$l^1_* = \{x \in l^1 : \sum_{i=1}^{\infty} i|x_i| < \infty\}.$$

**Proposition 5.36.** *The space $l^1_*$ is a normed subspace.*

*Proof.* The definition makes sense because for all $x \in l^1_*$, $|x_j| \leq j|x_j|$, so by comparison $\sum_{j=1}^{\infty} |x_j|$ converges, or in other words $x \in l^1$. If $x$ and $y \in l^1_*$ and $\alpha \in \mathbf{R}$ then $\alpha x + y \in l^1_*$ by the Algebra of Limits. Thus $l^1_*$ is a linear space. It is a normed vector space with the $l^1$-norm. □

Unfortunately, $l^1_*$ is not closed, and hence it is not a Banach space either. As an example, we take the sequence $x^{(n)}$ of elements in $l^1_*$ given by $x^{(n)}_j = j^{-(2+1/n)}$ for each $j \in \mathbf{N}^+$. The $l^1$-limit of $x^{(n)}$ is $(1/j^2)$ which lies outside of the subspace $l^1_*$. It is important therefore that we manually check that the limits of the sequences, with which we work, lie in $l^1_*$.

In order to show that we can work with the restriction of the mean operator $T^{(N)}$ onto $l^1_*$, we first observe that our initial configuration of individuals

$$Z(0) = (x_1, \ldots x_n) = x,$$

where $x_i$ is the number of individuals of family size $i$ $(1 \leq i \leq n,)$, and crucially $n$ is some finite natural number. In this case $Z(0) \in l^1_*$ because

$$\sum_{i=1}^{\infty} ix_i = \sum_{i=1}^{n} ix_i < \infty.$$

Thus we can work in the subspace $l^1_*$ instead of the whole $l^1$ provided we also have that $T^{(N)} l^1_* \subseteq l^1_*$.

**Proposition 5.37.** *The space $l^1_*$ is closed under the mean operator, that is, $T^{(N)}|_{l^1_*} : l^1_* \to l^1_*$.*

*Proof.* Suppose that $x \in l^1_*$. Then we obtain that

$$\begin{aligned}
\sum_{j=1}^{\infty} j|(T^{(N)}x)_j| &= \sum_{j=1}^{\infty} j|\sum_{i=1}^{\infty} q_{ij}x_i| \\
&\leq \sum_{i=1}^{\infty} |x_i| \sum_{j=1}^{\infty} j^2 p_j(w_i); \text{ Taylor expansion about 1} \\
&= \sum_{i=1}^{\infty} |x_i| \sum_{j=1}^{\infty} j^2 (p_j(1) + \frac{\epsilon_i}{N} p_j'(1) + O_j(N^{-2}) \\
&\leq \left( \sum_{j=1}^{\infty} j^2 p_j(1) + O(N^{-2}) \right) \|x\|_1 + \sum_{j=1}^{\infty} j^2 p_j'(1) \cdot \frac{\sup_i |\alpha_i|}{N} \cdot \sum_{i=1}^{\infty} i|x_i| \\
&= \sum_{j=1}^{\infty} j^2 p_j(1) \|x\|_1 + \frac{\|\alpha\|_\infty}{N} \sum_{j=1}^{\infty} j^2 p_j'(1) \sum_{i=1}^{\infty} i|x_i| + O(N^{-2}) \\
&< \infty,
\end{aligned}$$

because $x \in l^1_* \subseteq l^1$ and by the assumption on the distribution $p$ that the function $\sum_{j=1}^{\infty} j^2 p_j(w)$ is twice differentiable. Thus $T^{(N)}x \in l^1_*$. □

The restriction inherits the properties of the mean operator.

**Proposition 5.38.** *The operator $T^{(N)}|_{l^1_*}$ is bounded linear and compact.*

The proofs from §5.2 carry over. When we substitute in $i\alpha$ for $\epsilon_i$, the sums involved converge by the definition of the space $l^1_*$. Because of compactness, the spectral radius of $T^{(N)}|_{l^1_*}$ is an eigenvalue.

**Proposition 5.39** (Spectral radius)**.** *The spectral radius of $T^{(N)}|_{l^1_*}$ equals $1+\beta/N+O(1/N^2)$, where*

$$\beta = \sum_{j=1}^{\infty} j\epsilon_j p_j(1) = \sum_{j=1}^{\infty} j^2 \alpha_j p_j(1).$$

*Proof.* The spectral radius of a restricted operator is not greater than the spectral radius of the original operator:

$$r_\sigma(T^{(N)}|_{l^1_*}) \leq r_\sigma(T^{(N)}).$$

But, as $\sum_{j=1}^{\infty} j^2 p_j(w) < \infty$ and twice differentiable, with $\gamma \in l^\infty$, the sequences

$$v^{(N)} = \left( j p_j(1) + \frac{p_j(1)\gamma_j}{N} + O_j(N^{-2}) \right)_{j \geq 1}$$

lie in $l^1_*$, because of Taylor's Theorem with a second order error term. In Proposition 5.18 we showed that $v^{(N)}$ is an eigensequence of $T^{(N)}$ with eigenvalue $1+\beta/N+O(1/N^2)$. Thus it is an eigenvalue of the restriction too, and hence

$$r_\sigma(T^{(N)}|_{l^1_*}) \geq 1 + \frac{\beta}{N} + O(N^{-2}).$$

Summarising, we have

$$1 + \frac{\beta}{N} + O(N^{-2}) \leq r_\sigma(T^{(N)}|_{l^1_*}) \leq r_\sigma(T^{(N)}) = 1 + \frac{\beta}{N} + O(N^{-2}).$$

We therefore obtain that $r_\sigma(T^{(N)}|_{l^1_*}) = 1 + \beta/N + O(1/N^2)$. □

We illustrate the proposition with an example.

**Example 5.40** (Poisson offspring with increasing weight sequence)**.** Let the offspring number $X_i \sim Pois(w_i)$ depending on the parent's family size $i$. Then $p_j(w_i) = w_i^j e^{-1}/j!$. Using the unbounded weight sequence $w = (1+i\alpha/N)$ with some positive real number $\alpha$, and applying Proposition 5.39, we find that the spectral radius of the mean operator associated with the branching process is $1+\beta/N+O(N/2)$,

where

$$\begin{aligned}\beta &= \sum_{i=1}^{\infty} i\alpha p_i(1) = \sum_{i=1}^{\infty} \frac{i^2\alpha}{i!}e^{-1} \\ &= \alpha \sum_{j=0}^{\infty} \frac{j+1}{j!}e^{-1} = 2\alpha.\end{aligned}$$

In turn, the limits of the corresponding eigenvectors are scalar multiples of the sequence

$$v_i = \lim_{N\to\infty} v_i^{(N)} = ip_i(1) = \frac{\alpha e^{-1}}{(i-1)!}.$$

We close this chapter by giving an analogous result on the law of the limiting re-scaled branching process $\lim_{N\to\infty} N^{-1}Z^{(N)}(\lfloor Nt \rfloor)$ to Theorem 5.28. It is not clear whether the paths of the limiting process of the sequence of rescaled branching processes remain continuous if we exchange the weights $w_i = 1+\epsilon_i/N$ (where $\epsilon \in l^\infty$) with strictly increasing weights $1 + i\alpha_i/N$ (where $\alpha \in l^\infty$). In order to be able to handle this case, we must modify the re-scaled process slightly.

**Definition 5.41** (Re-scaled Galton-Watson process for CTF with weight sequence $w = (1 + i\alpha/N)$)**.** *We define the $N$th stage of the scaled process $X^{(N,\epsilon)}$ by*

$$X^{(N,\epsilon)}(t) = \frac{1}{N}Z^{(N,\epsilon)}(\lfloor Nt \rfloor),$$

*where the process $Z^{(N,\epsilon)}$ is an infinite-dimensional Galton-Watson process described in Definition 5.2 with weight sequence $w = (w_i)$ given for all $i \geq 1$ by*

$$w_i = \begin{cases} 1 + \frac{i}{N}\alpha, & \text{if } i \leq N \\ 1 + \alpha, & \text{otherwise.} \end{cases}$$

One could say that we are deceiving ourselves here because we have just made the weight sequence $w$ bounded. Nevertheless, the above branching process is analogous to the constant population size model studied in Part II. There, we used weight sequences of the form $w = (j^\alpha)$ but ultimately these sequences are bounded too because of the finite population size $N$. The largest possible reproductive weight that an individual could have was $N^\alpha$.

**Theorem 5.42.** *Let $X^{(N)}$ be a re-scaled G-W process for CTF with weight sequence $w = (1 + i\alpha/N)$. Assume Condition 5.24 on the second and third moments of the offspring distributions $p(w_i) = (p_j(w_i))$ and that the distribution of $X^{(N)}(0)$ tends to $\mu$, a probability measure on the space $l^1$. Let $\mu_1$ be a probability measure on $l^1$ given by*

$$\mu_1(A) = \mu\{x \in l^1 : \forall i\, x_i \in \mathbf{R}^+, \sum_{i=1}^{\infty} x_i \in A\}$$

*for all set $A \in \mathbf{B}(\mathbf{R}^+)$. Write $\zeta^{(N)}$ for the total population size in the $N$th stage of refinement defined by $\zeta^{(N)} = \sum_{i=1}^{\infty} X_i^{(N)}$.*

*Then $\zeta^{(N)}$ converges weakly to the unique in law diffusion $\zeta$ with initial distribution $\mu_1$ and generator $\tilde{L}$ given by*

$$(\tilde{L}f)(x) = \beta x f'(x) + \frac{\sigma^2}{2} f''(x)$$

*for all functions $f \in C_K^\infty(\mathbf{R})$, where $\sigma^2 = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \sigma_{ij}$, $\sigma_{ij} = \lim_{N\to\infty} \sigma_{ij}^{(N)}(l)$ for all $l \in \mathbf{N}^+$, $\sigma_{ij}^{(N)}(l)$ denoting the covariance of the offspring of an individual of family size $l$.*

Thus, even in the case of a strictly increasing weight sequence, total population of the rescaled branching process follows the Feller diffusion found in §5.3. Before moving on to the proof of the theorem, we collect the relevant definitions and

calculations given earlier in the chapter.

We define the mean operator at the $N$th level of refinement $T^{(N)} : l^1_* \to l^1_*$ for all $x \in l^1_*$, by

$$(T^{(N)}x)_j = \sum_{i=1}^{\infty} q^{(N)}_{ij} x_i,$$

where the kernel $q^{(N)}$ is given by

$$q^{(N)}_{ij} = jp_j(1) + jp'_j(1)\frac{\epsilon^{(N)}_i}{N} + O(N^{-2}),$$

$(p_i(1))$ being the probability mass function of the offspring number distribution of mean 1. As discussed above, here the weights are now of the form $w_i = 1 + \epsilon^{(N)}_i/N$, with

$$\epsilon^{(N)}_j = \begin{cases} j\alpha, & j < N \\ N\alpha, & \text{otherwise.} \end{cases}$$

We also define the operators $T$ and $C^{(N)} : l^1_* \to l^1_*$ by

$$(Tx)_j = \sum_{i=1}^{\infty} jp_j(1)x_i = jp_j(1)\sum_{i=1}^{\infty} x_i, \text{ and}$$
$$(C^{(N)}x)_j = \sum_{i=1}^{\infty} jp'_j(1)\epsilon^{(N)}_i x_i = jp'_j(1)\sum_{i=1}^{\infty} \epsilon^{(N)}_i x_i.$$

We immediately note that, following the argument in Proposition 5.37, $l^1_*$ is closed under the operator $T^{(N)}$ just as it is closed under $T$. The subspace is also closed under $C^{(N)}$ as $(jp'_j(1)) \in l^1_*$ according to our assumption of the parametrised distribution $p$.

Lastly, we write $C$ for the point-wise limit of the operator $C^{(N)}$: for all $x \in l^1_*$

we define the operator $C : l^1_* \to l^1_*$ by

$$(Cx)_j = \lim_{N\to\infty} (Cx)_j = jp'_j(1) \sum_{i=1}^{\infty} i\alpha x_i.$$

The operator is well-defined because series of the form $\sum_{i=1}^{\infty} ix_i$ converge.

We also recall that the eigenvectors corresponding to the spectral radius of the operator $T^{(N)}$ and its adjoint $T^{(N)\times}$ are scalar multiples of

$$v^{(N)} = \left( jp_j(1) + \frac{\gamma_j p_j(1)}{N} + O_j(N^{-2}) \right) \text{ and}$$
$$u^{(N)} = \left( 1 + \frac{\delta_j}{N} + O_j(N^{-2}) \right),$$

respectively.

**Definition 5.43** (The product operator $S$)**.** *We define the operator $S : l^1_* \to l^1_*$ by*

$$(Sx)_j = \sum_{i=1}^{\infty} v_j u_i x_i = v_j \sum_{i=1}^{\infty} u_i x_i = jp_j(1) \sum_{i=1}^{\infty} x_i,$$

*for all $x \in l^1_*$ and $j \in \mathbf{N}^+$.*

We note that, because of the special structure of our branching process, $S = T$. The finite dimensional version of $S$ would correspond to the product of the eigenvectors: the matrix $v\,u^T$. As in the finite dimensional case $S$ has the following properties that we shall exploit in the proof of the theorem.

**Proposition 5.44.** *The functional equation $S(T-I) = 0$ holds, where the operators $I$ and 0 denote the identity and zero operators. We also have that $(I-S)X^{(N)}(t) \to 0$ in probability uniformly on compact time intervals.*

*Proof.* In order to verify the first claim, we observe that, for all $x \in l^1_*$ and all $j \geq 1$,

we have

$$\begin{aligned}(STx)_j &= \sum_{i=1}^{\infty} v_j u_i \, ip_i(1) \sum_{k=1}^{\infty} x_k \\ &= \sum_{i=1}^{\infty} jp_j(1) \, ip_i(1) \sum_{k=1}^{\infty} x_k \\ &= jp_j(1) \sum_{k=1}^{\infty} x_k = (Sx)_j.\end{aligned}$$

For the second claim, we note that, for all $t \in \mathbf{N}^+$ and all initial configurations $x \in l^1_*$,

$$\begin{aligned}\mathrm{E}\left[X^{(N)}(t) - SX^{(N)}(t) \,|\, X^{(N)}(0) = Nx\right] &= \frac{1}{N} T^{\lfloor Nt \rfloor}(I - S)Nx \\ &= T^{\lfloor Nt \rfloor}(I - S)x \to 0,\end{aligned}$$

as $N \to \infty$ by a similar calculation of $(TSx)_j$. Hence we also have that $(I - S)X^{(N)} \to 0$ in probability. □

We are now ready to give a proof of the theorem. We mimic Amarjit Budhiraja and Dominik Reinhold's demonstration of Theorem 5.28, which is the finite dimensional version that we applied in §5.3. Budhiraja and Reinhold's proof is, in turn, based on the proof of Theorem 4.3.1 in Joffe and Metivier (1986):

**Theorem 5.45** (Joffe-Metivier)**.** *Let $Z_n$ be a multi-type critical positive regular Galton-Watson process with mean matrix $M$ and covariance matrices $\sigma_{i,j}$. Let $R$ be the projection matrix in the Perron-Frobenius decomposition of $M$. Then if $X^{(n)}(t) = \epsilon_n Z^n_{[nt]}$ with $X^{(n)}(0) = x_0^{(n)}$ converging to $x_0 \neq 0$, $n\epsilon_n \to \alpha \in \mathbf{R}^+$, the sequence*

$$\xi^{(n)}(t) = X^{(n)}(t) - n \int_0^t X^{(n)}(s-)(M - I) dA_n(s)$$

*will converge weakly to the unique diffusion in the half space* $xR \leq 0$ *with generator* $L$ *given by*

$$Lf = \alpha XR\frac{1}{2}\sum_{ij}\sigma_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j} \qquad \textit{for all } f\ C^2\textit{-functions on } xR \geq 0$$

*and initial condition* $x_0$.

Thus Budhiraja and Reinhold's contribution was the extension to near-critical branching processes, while ours is merely the adoption of the latter to the infinite dimensional space $l^1_*$.

*Proof of Theorem 5.42.* We define the auxiliary process $Y^{(N)}(t)$ by

$$Y^{(N)}(t) = y^{(N)} + N\int_0^t (T-I)X^{(N)}(s-)dA_N(s),$$

where the integrator $A_N = \lfloor Nt \rfloor/N$ and $y^{(N)}$ is any sequence in $l^1_*$: $Y^{(N)}(0) = y^{(N)}$.

Then the pair $(\bar{X}^{(N)}(k), \bar{Y}^{(N)}(k)) := (X^{(N)}(k/N), Y^{(N)}(k/N))$ is a Markov chain with one-step transition probabilities

$$\bar{P}^{(N)}((x,y),(x',y')) = \bar{Q}^{(N)}(x,x')\chi_{y'=y+(T-I)x},$$

where $\bar{Q}^{(N)}(\cdot,\cdot)$ is the transition probability of the process $\bar{X}^{(N)}$ defined above.

Finally we define the generators $L^{(N)}$ and $\bar{L}^{(N)}$ by

$$\begin{aligned}(\bar{L}^{(N)}f)(x,y) &= \sum_{x',y' \in l^1_*} \bar{P}^{(N)}((x,y),(x',y'))\,[f(x',y') - f(x,y)] \text{ and} \\ L^{(N)} &= N\bar{L}^{(N)},\end{aligned}$$

for all functions $f \in C^2(l^1_* \times l^1_*)$ and $x$, $y \in l^1_*$. Then we may use Dynkin's formula

to deduce that, for each $f \in C^2(l^1_* \times l^1_*)$, the process

$$M^{(N)}(t) := f((X^{(N)})(t), Y^{(N)}(t)) - \int_0^t (L^N f)(X^{(N)})(s-), y^{(N)}(s-))dA_N(s)$$

is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma(\{(X^{(N)}(s), Y^{(N)}(s)) : 0 \leq s \leq t\})$.

Using test functions of the form $f(x,y) = \phi(x-y)$, where $\phi \in C^\infty_K(l^1_*)$, and setting $y^{(N)} = 0$ for all $N \in \mathbf{N}^+$, Taylor's Theorem gives us that the quantity

$$\begin{aligned}
&\phi(X^{(N)}(t) - Y^{(N)}(t)) - \phi(X^{(N)}(0)) - \\
&- \int_0^t \sum_{i=1}^{\infty} \left[(T^{(N,\epsilon)}) - T)X^{(N)}(s-)\right]_i \frac{\partial \phi}{\partial x_i}(X^{(N)}(s-) - Y^{(N)}(s-))\, dA_N(s) - \\
&- \frac{1}{2}\int_0^t \sum_{i,j,l=1}^{\infty} \sigma^{(N)}_{ij}(l)(X^{(N)}(s-))_l \frac{\partial^2 \phi}{\partial x_i \partial x_j}(X^{(N)}(s-) - Y^{(N)}(s-))\, dA_N(s) \\
&+ E_N(t)
\end{aligned}$$

is a martingale where the remainder term $E_N(t)$ is of the form

$$E_N(t) = \int_0^t R_N(X^{(N)}(s-), Y^{(N)}(s-))\, dA_N(s)$$

with $R_N = \sum_{x \in l^1_*} R^{(z)}_N$ where $R^{(z)}_N$ is defined for all $x$ and $y \in l^1_*$ by

$$\begin{aligned}
R^{(z)}_N(x,y) = \frac{1}{2}\sum_{i,j=1}^{\infty} &\left(\frac{z}{N} - Tx\right)_i \left(\frac{z}{N} - Tx\right)_j \cdot \\
&\cdot P(Nx, z)\left[\frac{\partial^2 \phi}{\partial x_i \partial x_j}(\eta(z) - \frac{\partial^2 \phi}{\partial x_i \partial x_j}(x-y)\right].
\end{aligned}$$

Here $\eta(z)$ is an element in the space $l^1_*$ such that $\|\eta(z) - (x-y)\| \leq \|\frac{z}{N} - Tx\|$. From our assumption that the third moments of the offspring numbers for all family sizes

are uniformly bounded in $N$ (Condition 5.24), following Joffe and Metivier (1986), we obtain that

$$\sup_{t\in[0,a]} |E_N(t)| \to 0 \text{ as } N \to \infty$$

in probability for all closed intervals $[0, a]$, with $a \in \mathbf{R}^+$.

Next, with the help of Proposition 5.44 we will show that

$$\lim_{N\to\infty} \int_0^t \mathrm{E}^{Nx} |(\tilde{L}^{(N)}\phi)(X^{(N)}(s-), \xi^{(N)}(s-)) - (L\phi)(\xi^{(N)}(s-))|\, dA_N(s) = 0,$$

where the quantity $\xi^{(N)} = X^{(N)} - Y^{(N)}$ and the differential operators $\tilde{L}^{(N)}$ and $L$ are defined by

$$(\tilde{L}^{(N)}f)(x,y) = \sum_{i=1}^{\infty} ((T^{(N)} - T)x)_i \frac{\partial f}{\partial y_i}(y) + \frac{1}{2} \sum_{i,j,l=1}^{\infty} x_l \sigma_{ij}(l) \frac{\partial^2 f}{\partial y_i \partial y_j}(y)$$

$$(Lf)(y) = \sum_{i=1}^{\infty} (CSy)_i \frac{\partial f}{\partial y_i}(y) + \frac{1}{2} \sum_{i,j,l=1}^{\infty} (Sy)_l \sigma_{ij}(l) \frac{\partial^2 f}{\partial y_i \partial y_j}(y),$$

respectively, for all $f \in C_K^\infty(l_*^1)$ and $x,\ y \in l_*^1$. We kept the notation compact by writing $\mathrm{E}^{Nx}$ to denote the conditional expectation on $\{X^{(N)}(0) = Nx\}$. In order to

prove this claim, we calculate

$$
\begin{aligned}
&(\tilde{L}^{(N)}\phi)(X^{(N)}(s-),\xi^{(N)}(s-)) - (L\phi)\xi^{(N)}(s-) = \\
&= \sum_{i=1}^{\infty} \left[(C^{(N)}X^{(N)}(s-))_i - (CS(X^{(N)}(s-) - Y^{(N)}(s-))_i\right] \frac{\partial\phi}{\partial z_i}(X^{(N)}(s-) - Y^{(N)}(s-)) + \\
&\quad + \frac{1}{2}\sum_{i,j,l=1}^{\infty} \left[X_l^{(N)}(s-) - (S(X^{(N)}(s-) - Y^{(N)}(s-)))_l\right] \sigma_{ij} \frac{\partial^2\phi}{\partial z_i \partial z_j}(X^{(N)}(s-) - Y^{(N)}(s-)); \\
&\quad \text{because } S(T-I) = 0 \\
&= \sum_{i=1}^{\infty}((C^{(N)} - CS)X^{(N)}(s-))_i \frac{\partial\phi}{\partial z_i}(X^{(N)}(s-) - Y^{(N)}(s-)) + \\
&\quad + \frac{1}{2}\sum_{i,j,l=1}^{\infty} ((I-S)X^{(N)}(s-))_i \sigma_{ij}(l) \frac{\partial^2\phi}{\partial z_i \partial z_j}(X^{(N)}(s-) - Y^{(N)}(s-)).
\end{aligned}
$$

Thus, because $\lim_{N\to\infty} \mathrm{E}^{Nx}(I-S)X^{(N)}(s-) = 0$ by Proposition 5.44 and $\lim_{N\to\infty} C^{(N)} = C$, letting $N \to \infty$, we obtain that

$$\mathrm{E}^{Nx}[(\tilde{L}^{(N)}\phi)(X^{(N)}(s-),\xi^{(N)}(s-)) - (L\phi)(\xi^{(N)}(s-)] \to 0.$$

Finally, since $A_N(t) = \frac{\lfloor Nt \rfloor}{N} \to t$, as $N \to \infty$, applying the Monotone Convergence Theorem, we conclude that

$$
\begin{aligned}
&\lim_{N\to\infty} \int_0^t \mathrm{E}^{Nx}|(\tilde{L}^{(N)}\phi)(X^{(N)}(s-),\xi^{(N)}(s-)) - (L\phi)(\xi^{(N)}(s-))|\, dA_N(s) = \\
&= \int_0^t \lim_{N\to\infty} \mathrm{E}^{Nx}|(\tilde{L}^{(N)}\phi)(X^{(N)}(s-),\xi^{(N)}(s-)) - (L\phi)(\xi^{(N)}(s-))|\, dA_N(s) \\
&= 0.
\end{aligned}
$$

So far we have merely identified the possible limit point of the process' distributions. We must also show that $\xi$ is a tight sequence. Granted tightness, any subsequence of the distributions will have a convergent subsequence: exactly what

we need to show that our candidate is really a limit. One proves the tightness of $X^{(N)}$ by appealing to the Aldous-Rebolledo Criterion in Joffe and Metivier (1986). One finds the relevant calculations for the finitely-many-types near-critical branching process in the Application section of Joffe and Metivier (1986). These carry over to our case, (and we will not replicate them here), but in this infinite dimensional set-up these conditions themselves cannot guarantee tightness. Mass can also escape to infinity along the sequence of types. Fortunately this does not occur, because our types are family size classes which, in turn, are always bounded by the total population size. By calculating the largest (in modulus) eigenvalue, $1+\beta/N+O(1/N^2)$, of the mean operator, we have demonstrated that the population size does not explode when time runs in units of $N$. The expectation grows linearly; so the probability of infinite population size on compact time intervals is zero.

As $X^{(N)}$ is a tight sequence in $D(\mathbf{R}^+, l^1_*)$, so is $Y^{(N)}$ by Proposition 5.44. Thus $\xi^{(N)} = X^{(N)} - Y^{(N)}$ is tight too. It follows now from the above calculation that if $\xi$ is any weak limit of the sequence of processes $\xi^{(N)}$, then the quantity

$$\phi(\xi(t)) - \phi(\xi(0)) - \int_0^t (L\phi)(\xi(s))\, ds$$

is a $\sigma(\{\xi(s) : 0 \le s \le t\})$-adapted martingale. Therefore $\xi^{(N)}$ weakly converges to a process $\xi$ with infinitesimal generator $L$ and initial condition $x$. We note that the initial condition for $Y^{(N)}$ was chosen to be $y^{(N)} = 0$.

Because $u = (1, 1, 1, \dots)$ is an eigensequence of $T^\times$,

$$\sum_{i=1}^{\infty} (T - I)x_i = \sum_{i=1}^{\infty} u_i (T - I)x_i = 0$$

and hence we obtain that

$$\zeta^{(N)} = \sum_{i=1}^{\infty} X_i^{(N)} = \sum_{i=1}^{\infty} u_i X_i^{(N)} = \sum_{i=1}^{\infty} u_i (X_i^{(N)} - Y_i^{(N)}) = \sum_{i=1}^{\infty} u_i \xi_i^{(N)}.$$

Now it follows from the convergence of $\xi^{(N)}$ that the re-scaled total population

$$\zeta^{(N)}(t) \to \sum_{i=1}^{\infty} \xi_i u_i = \sum_{i=1}^{\infty} (\lim_{N\to\infty} N^{-1} Z(\lfloor Nt \rfloor))_i =: \zeta$$

in distribution, as $N \to \infty$, for all $t \in \mathbf{R}^+$.

It remains to recover the generator of the limit process $\zeta$ from $L$, the infinitesimal generator of $\xi$.

To this end, we fix $f \in C_K^\infty(\mathbf{R}^+)$ and define $\psi \in C_K^\infty(l_*^1)$ by $\psi(z) = g(\sum_{i=1}^\infty z_i)$ for all $z \in l_*^1$. Applying $L$ on $\psi(z)$ now yields

$$\begin{aligned}(Lg)(\sum_{i=1}^{\infty} z_i) = (L\psi)(z) = \\ = \sum_{i=1}^{\infty} (CSz)_i \frac{\partial \psi}{\partial z_i}(z) + \frac{1}{2} \sum_{i,j,l=1}^{\infty} (Sz)_l \sigma_{ij}(l) \frac{\partial^2 \psi}{\partial z_i \partial z_j}(z).\end{aligned}$$

Thus we need to calculate the following quantities.

$$\begin{aligned}\frac{\partial \psi}{\partial z_i}(z) &= \frac{\partial}{\partial z_i} g\Big(\sum_{k=1}^{\infty} z_k\Big) = g'\Big(\sum_{i=1}^{\infty} z_k\Big) \cdot 1, \\ \frac{\partial \psi}{\partial z_i \partial z_j}(z) &= \frac{\partial^2}{\partial z_i \partial z_j} g\Big(\sum_{k=1}^{\infty} z_k\Big) = g''\Big(\sum_{i=1}^{\infty} z_k\Big) \cdot 1 \cdot 1,\end{aligned}$$

for all $i$ and $j \geq 1$; together with

$$\sum_{i=1}^{\infty} (CSz)_i = \sum_{i=1}^{\infty} i p_1'(1) \sum_{j=1}^{\infty} j \epsilon_j p_j(1) \cdot \sum_{k=1}^{\infty} z_k = 1 \cdot \beta \cdot \sum_{k=1}^{\infty} z_k;$$

and finally

$$\sum_{i,j,l=1}^{\infty}(Sz)_l\sigma_{ij}(l) = \sum_{i,j,l=1}^{\infty} lp_l(1)\sum_{k=1}^{\infty} z_k\sigma_{ij};\ \text{as } \sigma_{ij}(l) = \sigma_{ij} \text{ for all } l \in \mathbf{N}^+$$
$$= \sum_{i,j=1}^{\infty}\sigma_{ij}\sum_{k=1}^{\infty} z_k = \sigma^2\sum_{k=1}^{\infty} z_k.$$

We therefore obtain that $\zeta$ is a Feller diffusion with generator

$$(\tilde{L}g)(x) = \beta x g'(x) + \frac{\sigma^2}{2} x g''(x),$$

for all $x \in \mathbf{R}^+$ and $g \in C_K^\infty(\mathbf{R}^+)$, completing the proof. □

**Remark 5.46.** The question arises whether more rapidly increasing weight sequences, such as $w = (1 + j^2/N)$, still leave us with a diffusion approximation. In the important case of $Pois(w)$ distributed offspring numbers, it is clear that Condition 5.24 fails because $\sup_i w_i = 1 + N$, leading to discontinuous paths of the approximating process.

## 5.5 Conclusion

The constrain of a constant total population size throughout the generations aims to capture competition in mating. However, as we have seen in §§1.2 and 2.1 among the Hutterites in the U.S.A. and among the inhabitants of the Saguenay-Lac-Saint-Jean region in Québec, Canada, this competition is negligible. Thus the development of these rapidly growing populations is better described by a branching process rather than by a constant population size model.

As in the constant population size model of §2.4, cultural transmission of fitness is captured by assigning reproductive weights to individuals according to their

family size. The individuals' average offspring numbers equal the weights that they receive. Unlike in the constant population size model, here the individuals reproduce independently from each other.

Multi-type branching processes with several types, like ours, are difficult to analyse. We made the model more tractable by choosing the reproductive weights close to 1, making the branching process near-critical. In our calculation of the key quantity, the principal eigenvalue of the mean matrix belonging to the branching process, we were able to exploit the special structure of the matrix.

Using the principal eigenvalue, in §5.3, we approximated the Galton-Watson process by a diffusion, and studied the behaviour of the process when it is conditioned to non-extinction, using the results of Budhiraja and Reinhold (2010).

Our approach also works for a CTF Galton-Watson process with countably many family sizes. In §5.4 we showed with the help of Joffe and Metivier (1986) that the diffusion approximation works even when the weight sequence $w = (1 + j\alpha/N)$ (a strictly increasing sequence). In the end, we also saw that the sequence $w = (1 + j^2/N)$ increases too fast: the diffusion approximation breaks down.

# 6 Part IV: A branching diffusion with unlimited different family size types

Part I convinced us that, as long as the reproductive weights are bounded, the limiting coalescent of the rescaled population processes is a time-changed $n$-coalescent process. The rôle of this short part is to demonstrate that an unbounded weight sequence, increasing in family size, can potentially give a limiting coalescence process with multiple mergers. Often, continuous-state models are much easier to analyse than processes with discrete state space. This is so because continuous stochastic processes confer valuable regularity of paths, while our previous Markov chains could jump from any state to any other state with a positive probability.

In this model the expected offspring number of a mother diffuse on the positive half-line with negative drift. Prior to their death at the origin the individuals give birth to children at an exponential rate of 1. The offspring is born at the mother's current location. In §6.1 our aim is to find a pair of drift and diffusion coefficients which make it possible to interpret the position in space as family size. An individual born at $x \in \mathbf{R}^+$ shall have $x$ offspring on average. As a mother diffuses around her mean offspring number, her children born early will more or less inherit her family size. Because mothers slowly drift to 0, children born very late do not inherit their mother's family size, instead they have less offspring themselves because they receive less attention from their old mother. (The elderly mother may have many children by their birth, who have depleted her resources.) Berestycki et al. (2013) showed that the genealogies of a branching diffusion absorbed at the origin with underlying motion given by the SDE

$$dY_t = -\sqrt{2}dt + dB_t; Y_0 = y$$

roughly speaking evolve according to the Bolthausen-Sznitman coalescent. This statement is not precise, as the branching diffusion with the underlying spatial motion given above is critical, and hence it dies out. The drift coefficient of $Y_t$ needs to be slightly greater than $-\sqrt{2}$. This process exhibits multiple mergers of lineages. In §6.2 we rescale time to fit our model to (Berestycki et al., 2013)'s result.

## 6.1 Derivation of the underling spatial motion

Recall that individuals move according to a diffusion on $\mathbf{R}^+$ and are absorbed as they hit the origin. Define $\tau = \inf\{t > 0 : X_t = 0\}$ to be the time of absorption of an individual born at $x \in \mathbf{R}^+$ whose position $X_t$, at time $t$, is given by the stochastic differential equation (abbreviated SDE)

$$\begin{aligned} dX_t &= \mu(X_t)dt + \sigma(X_t)dB_t, \text{ for all } t > 0 \text{ and} \\ X_0 &= x. \end{aligned}$$

Then $\tau$ is a stopping time with respect to the natural filtration $(\mathcal{F}_t) := \sigma\{B_s : s \leq t\}$. Our objective is to choose $\mu(x)$ and $\sigma(x)$ in the SDE above so that

(i) a strong solution exists on $(0, \infty)$,

(ii) the stopping time $\tau \in L^1$ and

(iii) $\mathrm{E}^x \tau = x$.

It turns out that we can assume that $\mu$ and $\sigma$ are constants. Evidently, however, $\mu < 0$ otherwise (ii) fails. In this simple case, one can explicitly find the distribution of the lifespan $\tau$. We do this by slightly modifying the calculation of the Laplace transform of Brownian motion with drift stared from 0 found in the first chapter of

Rogers and Williams (2000).

**Proposition 6.1.** *Let $X_t = x + \mu t + \sigma B_t$ be a Brownian motion with drift started at $x \in \mathbf{R}^+$, and define the hitting time $\tau$ as above. Then the distribution of $\tau$ is given by its Laplace transform: $\mathrm{E}^x e^{-\lambda\tau} = e^{-\alpha(\lambda)x}$, where $\alpha(\lambda) = -(\sqrt{\mu^2 + 2\sigma^2\lambda} + \mu)/\sigma^2$ for all $\lambda > 0$.*

*Proof.* Choose $\lambda > 0$. As $\exp(\theta B_t - \frac{1}{2}\theta^2 t)$ is a martingale with respect to the natural filtration $(\mathcal{F}_t)$ for all $\theta \in \mathbf{R}$,

$$\begin{aligned} e^{-\theta x}\exp(\theta X_t - \lambda t) &= e^{-\theta x}\exp(\theta x + \mu\theta t + \sigma\theta B_t - \lambda t) \\ &= \exp[\sigma\theta B_t - (\lambda - \mu\theta)t] \end{aligned}$$

is a martingale if

$$\begin{aligned} &\lambda - \mu\theta = \frac{1}{2}(\sigma\theta)^2; \\ &\frac{1}{2}\sigma^2\theta^2 + \mu\theta - \lambda = 0; \\ &\theta_{1,2} = \frac{\pm\sqrt{\mu^2 + 2\sigma^2\lambda} - \mu}{\sigma^2}. \end{aligned}$$

Since $\lambda$ is positive and $\mu$ is negative, $\alpha := \theta_2 < 0 < \theta_1$; and hence the martingale $e^{-\alpha x}\exp(\alpha X_t - \lambda t)$ is bounded on the interval $[0, \tau]$. We can therefore apply the Optional Sampling Theorem to obtain that

$$\begin{aligned} 1 &= \mathrm{E}^x[e^{-\alpha x}\exp(\alpha X_0 - \lambda 0)] = \mathrm{E}^x[e^{-\alpha x}\exp(\alpha X_\tau - \lambda\tau)] \\ &= e^{-\alpha x}\mathrm{E}^x[\exp(-\lambda\tau)]. \end{aligned}$$

Thus the Laplace transform of $\tau$ is given by $\mathrm{E}^x e^{-\lambda\tau} = e^{\alpha(\lambda)x}$. We note that $\alpha$ is negative for all positive values of $\lambda$. □

Extending the work above, we calculate an individual's mean lifespan. Differentiating with respect to $\lambda$, and using the Chain Rule we obtain

$$\begin{aligned}(\mathrm{E}^{\,x}e^{-\lambda\tau})' &= \exp\left(-\frac{\mu+\sqrt{\mu^2+2\sigma^2\lambda}}{\sigma^2}x\right)\left(-\frac{1}{2}\frac{x}{\sigma^2\sqrt{\mu^2+2\sigma^2\lambda}}\right)2\sigma^2 \\ &= -\frac{x}{\sqrt{\mu^2+2\sigma^2\lambda}}e^{-\frac{\mu+\sqrt{\mu^2+2\sigma^2\lambda}}{\sigma^2}x} = -\frac{x}{\sqrt{\mu^2+2\sigma^2\lambda}}e^{\alpha(\lambda)}.\end{aligned}$$

Thus the expected lifespan is

$$\mathrm{E}^{\,x}\tau = -\lim_{\lambda\downarrow 0}\frac{d}{d\lambda}\mathrm{E}^{\,x}e^{-\lambda\tau} = \frac{x}{|\mu|}e^{-\frac{|\mu|+\mu}{\sigma^2}x} = -\frac{x}{\mu}.$$

Choosing $\mu = -1$ and $\sigma \in \mathbf{R}$, $X_t$, the Brownian motion with drift started at $x$, satisfies conditions (i), (ii) and (iii).

Reusing previous work and differentiating once again, we find the variance of the lifespan $\tau$ of an individual born at $x \in \mathbf{R}^+$.

$$\begin{aligned}\left(e^{\alpha(\lambda)x}\right)'' &= -\left(\frac{x}{\sqrt{\mu^2+2\sigma^2\lambda}}\right)' e^{\alpha(\lambda)x} + \left(\frac{x}{\sqrt{\mu^2+2\sigma^2\lambda}}\right)^2 e^{\alpha(\lambda)x} \\ &= \left(\frac{1}{2}\,\frac{x}{(\mu^2+2\sigma^2\lambda)^{\frac{3}{2}}}\,2\sigma^2 + \frac{x^2}{\mu^2+2\sigma^2\lambda}\right)e^{\alpha(\lambda)x} \\ &= \left(\frac{x^2}{\mu^2+2\sigma^2\lambda} + \frac{\sigma^2 x}{(\mu^2+2\sigma^2\lambda)^{\frac{3}{2}}}\right)e^{\alpha(\lambda)x}.\end{aligned}$$

Thus the second moment of the lifespan $\tau$ is

$$\mathrm{E}^{\,x}\tau^2 = \lim_{\lambda\downarrow 0}\frac{d^2}{d\lambda^2}\mathrm{E}^{\,x}e^{-\lambda\tau} = \left(\frac{x^2}{\mu^2}+\frac{\sigma^2 x}{|\mu|^3}\right)e^{\alpha(0)x} = \frac{x^2}{\mu^2}-\frac{\sigma^2 x}{\mu^3};$$

whence the variance of $\tau$ is

$$\operatorname{var}^x \tau = \mathrm{E}^x \tau^2 - (\mathrm{E}^x \tau)^2 = \frac{x^2}{\mu^2} - \frac{\sigma^2 x}{\mu^3} - \frac{x^2}{\mu^2} = -\frac{\sigma^2 x}{\mu^3}.$$

In particular, if we choose $\mu = -1$ and $\sigma = 1$, then $\mathrm{E}^x \tau = \operatorname{var}^x \tau = x$. As individuals produce offspring at exponential rate 1 throughout their life time, their expected number of offspring is

$$\mathrm{E}^x[Pois(1 \cdot \tau)] = \mathrm{E}^x[\mathrm{E}[Pois(\tau)|\tau]] = \mathrm{E}^x \tau = -\frac{x}{\mu};$$

and their variance is

$$\operatorname{var}^x[\mathrm{E}[Pois(1\tau)|\tau]] = \mathrm{E}[\tau^2 + \tau] - \mathrm{E}[\tau]^2 = -\left(\frac{\sigma^2}{\mu^3} + \frac{1}{\mu}\right) x.$$

The conditional expectation is taken with respect to a Poisson process with rate parameter 1. Notice that in limit theorems for branching Markov processes only these two quantities play a rôle. It would be most welcome if the mean and the variance of the offspring number could be chosen to be equal, resembling a Poisson distribution. Under this model, this is not possible. A particularly nice choice for the parameter $\sigma$, however, is $\sigma = -\mu$. This gives offspring variance $-\frac{2}{\mu}x$, twice the offspring mean.

## 6.2 The coalescent dual of the model

As the introduction advertised, we apply the convergence theorem in (Berestycki et al., 2013) to determine the coalescent process associated with our branching diffusion model. To begin with, we describe a wide class of coalescent processes.

**Definition 6.2** ($\Lambda$-coalescent process)**.** *Assume that $\Lambda$ is a finite measure on the interval $[0,1]$. Whenever there remain $b$ lineages any $k$ of them may merge with equal instantaneous rate*

$$\lambda_{bk} = \int_0^1 x^{k-2}(1-x)^{b-k}\Lambda(dx).$$

We have already met the most important specimen of the $\Lambda$-coalescent processes in Part I: choosing $\Lambda$ to be the unit mass at the origin gives us the familiar $n$-coalescent. Another special case, the $U$-coalescent, is obtained by setting $\Lambda = \lambda|_{[0,1]}$ the Lebesgue measure on $[0,1]$. Thus the transition rates become

$$\begin{aligned}\lambda_{bk} &= \int_0^1 x^{k-2}(1-x)^{b-k}dx; \quad \text{as } \int_0^1 x^{\alpha-1}(1-z)^{\beta-1}dx = B(\alpha,\beta)\\ &= \frac{(k-2)!(b-k)!}{(b-1)!} = \frac{b}{k(k-1)}\Big/\binom{b}{k},\end{aligned}$$

for all $2 \leq k \leq b$. Hence a $U$-coalescent admits multiple mergers: several lineages can coagulate at the same time. The $U$-coalescent sometimes is also called the Bolthausen-Sznitman coalescent in honour of the two mathematicians: Erwin Bolthausen and Alain-Sol Sznitman.

Birkner et al. (2005) showed that the coalescent dual of any continuous-state branching process with an $\alpha$-stable branching mechanism is a time-changed $\beta$-coalescent ($\Lambda = B(2-\alpha,\alpha)$). In particular, the genealogy of a continuous-state branching process whose branching mechanism is of the form $\psi(u) = \alpha u + \beta u \log u$ evolves according to the $U$-coalescent.

We need some notation to introduce (Berestycki et al., 2013)'s results. Define the sequence $(a_N)$ by setting

$$a_N = \frac{2\pi}{(\log N + 3\log\log N)^2},$$

and consider a sequence of branching Brownian motions absorbed at the origin with $\mu_N = -\sqrt{2 - a_N}$ and $\sigma_N = 1$. So $X_N(t) = x + \mu_N t + B_t$. This is a sequence of supercritical processes approaching a critical branching process. The branching Brownian motion is critical with $\mu = -\sqrt{2}$ in the sense that it almost surely dies out if and only if $\mu \leq -\sqrt{2}$. Write $M_N(t)$ for the total population size at time $t$. We index the individuals according to decreasing distance from the origin: $X_{N,1}(t) \geq X_{N,2}(t) \geq \cdots \geq X_{N,M_N}(t)$.

As most particles stay below the barrier $L = \frac{1}{\sqrt{2}}(\log N + 3 \log\log N)$,

$$Z_N(t) = \sum_{i=1}^{M_N} \exp(\mu_N X_{N,i}(t)) \sin\left(\frac{\pi X_{N,i}(t)}{L}\right) \chi_{\{X_{N,i}(t)<L\}}$$

is a good measure of the total population size $M_N(t)$. $Z_N(t)$ predicts the number of individuals just after time $t$.

Berestycki et al. (2013) showed that whenever some conditions are met on the initial distribution of the individuals the coalescent processes of the above branching processes tend to the $U$-coalescent in finite dimensional distributions.

**Condition 6.3.** *Define* $V_N(t) = \frac{1}{N(\log N)^2} Z_N((\log N)^3 t)$. $V_N(0)$ *converges to some probability distribution on* $\mathbf{R}^+$.

**Condition 6.4.** *Define the random variables* $Y_N$ *by setting* $Y_N(t) = \sum_{i=1}^{M_N(t)} e^{\mu X_{i,N}(t)}$. *We require that* $Y_N(0)/N(\log N)^3 \to 0$ *as* $N \to \infty$.

**Theorem 6.5** (Berestycki et al. (2013))**.** *Assume the two conditions above. Then the rescaled sequence of population sizes* $\frac{1}{2\pi N} M_N((\log N)^3 t)$ *converges in finite-dimensional distributions to the continuous-state branching process with branching mechanism* $\psi(u) = \alpha u + 2\pi^2 u \log u$ *and initial distribution* $\nu$.

*Also, if* $\nu(\{0\}) = 0$, *then the coalescent processes of the branching Brownian mo-*

*tions* $(\Pi_N(s))_{0\le s\le 2\pi t}$ *converge in finite-dimensional distributions to the* $U$*-coalescent* $(\Pi(s))_{0\le s\le 2\pi t}$ *as* $N$ *becomes large, for any fixed* $t \in \mathbf{R}^+$.

For our interpretation of an individual's position in $\mathbf{R}^+$ as her family size, it is important that the underlying motion is $X_t = x + \mu t - \mu B_t$, (see the calculation above). We rescale time to obtain the required law $Y_t = y - \sqrt{2 - a_N}t + B_t$ for the limit theorem:

$$X_{\frac{t}{\mu^2}} - x = \mu\frac{t}{\mu^2} + \mu B_{\frac{t}{\mu^2}} \simeq \frac{1}{\mu}t + \tilde{B}_t$$

by Brownian scaling. We therefore need to choose $y = x$ and $\mu_N = -\frac{1}{\sqrt{2-a_N}}$. The coalescent duals of our sequence of branching Brownian motions therefore converge to $(\Pi(s/2))_{0\le s\le 2\pi t}$.

## 6.3 Concluding remarks

We have seen above that a time rescaled version of the branching diffusion in Berestycki et al. (2013) can be interpreted as a model for CTF. The associated coalescent process has multiple mergers, demonstrating that cultural transmission of fitness can lead to models with different qualitative features.

But we have a general remark on the applicability of Berestycki et al. (2013).

**Remark 6.6.** How large does the approximate population size $N$ have to be so that $\mu_N = -\sqrt{2 - a_N}$ is close to $-\sqrt{2}$? If

$$a_N = \frac{2\pi^2}{(\log N + 3\log\log N)^2} < \frac{2\pi^2}{(\log N)^2} < 10^{-4},$$

say, then $\exp(100\sqrt{2}\,\pi) < N$, that is, $N > 8.97\cdot 10^{192}$! Thus, so that the $U$-coalescent approximation describe well the genealogies of the population, we need an intergalac-

tic alien species. (The observable universe contains about $3 \cdot 10^{23}$ stars.) Given such a space-farer population, we can assume however that their society is sophisticated enough to exhibit CTF.

Finally, we should also note that, although the rescaled process exhibits multiple mergers, a direct comparison is not possible with the limit dual of our extended Wright-Fisher model. Generations have a different interpretation here as previously. In the extended Wright-Fisher model time is measured in generations, while in this branching diffusion model for CTF time is measured in years. Generations overlap.

# 7 Summary and future directions of research

In this short, final chapter, we sum up and outline the future directions of research. The reader finds more detailed conclusions at the end of each Part.

Cultural transmission of fitness is the phenomenon that the number of an individual's siblings influences the number of her own offspring. In §2.4 we proposed an extension of the classical Wright-Fisher model for haploid populations of constant population size $N$ to allow for setting the magnitude of these 'influences' for each family size separately. This was achieved by assigning a reproductive weight to each individual in the parental generation depending on their family size: an individual of family size $i$ receives the weight $w_i$. This individual is chosen as the mother of any given individual in the new generation with probability $w_i / \sum_{j=1}^{N} N p_j w_j$, where $p_j$ stands for the proportion of individuals of family size $j$ in the parental generation.

In §3.1 we showed that, in the infinite population size limit, the family size distribution is a convex combination of Poisson distributions. In §3.3 we studied the moments of and the correlation between the family size of a randomly picked individual from the population and of its parent's family size. We also expressed the parameters of the model in terms of these two distributions. In §3.4 we investigated the genealogies of a sample taken from the population. For our main result in Part I, we used a collapse of structure argument to show that, in fact, the ancestral process (time measured in units of the population size) converges weakly to a rescaled $n$-coalescent: a process in which only pairs of lineages coalesce as the population size tends to infinity.

**Theorem 7.1.** *Measuring time in units of the total population size $N$, the ancestral process $A_N(\lfloor Nt \rfloor)$ converges weakly to the rescaled $n$-coalescent $A(ct)$, as $N$ becomes*

*large. The scaling factor $c$ is given by*

$$c = \sum_{j=1}^{\infty} \frac{j-1}{p_j} \gamma_j^2,$$

*in terms of the time-stationary family size distribution $(p_j)$ and the invariant measure $(\gamma_j)$ of the backward migration process of the lineages.*

This theorem requires that the family size distributions $(p_j(t))$ converge in time. The dependence of this convergence on the choice of the weight sequence $w$ was investigated in §3.2. In §3.4 we described a method for calculating $\gamma$.

The $n$-coalescent is the limit coalescent dual of some of the most widely accepted forward-in-time models such as the Wright-Fisher, the Moran and some of the Cannings models. The fact that cultural transmission of fitness, choosing an appropriate effective coalescent population size, has a place in standard coalescent theory, should therefore come as a relief to many practitioners of population genetics, our result justifies their long-standing practice.

As Murphy (2007) pointed out, the 1987 U.S. National Survey shows that the trans-generational correlation in family size varies slightly from generation to generation which, in turn, suggests that the family size distribution $(p_j(t))$ is not time-stationary in reality. However, we saw in §3.1 that our recursion for $(p_j(t))_{t\geq 0}$ sometimes has a periodic solution. So our model might fit the U.S. population with a large period. For this reason, it is worth extending our coalescent approximation result above, so that it covers periodicity of $(p_j(t))$ as well. This might be done in a similar fashion as varying population size is treated in Gillespie (2004), for example.

In Part II, we presented results of numerical simulations of the above extended Wright-Fisher model with strictly increasing weight sequences of the form $w^{(\alpha)} = (j^\alpha)$, $\alpha > 0$. We observed a phase transition in family size frequency at the point

$\alpha = 1$; and we reported multiple mergers occurring in the model's coalescent dual. We also used the data to approximate the coalescent's natural time scale.

These numerics indicate that the model is intriguing from a mathematical point of view; and it is desirable to obtain analytical results. A strictly increasing weight sequence such as $w^{(\alpha)}$ seems to lack biological relevance, however.

A branching process had been proposed by Austerlitz and Heyer (1998) to capture CTF, but in Part III we worked with the branching process version of our extended Wright-Fisher model. Individuals are again classified by their family size but they reproduce independently; the offspring number of an individual of family size $i$ has distribution $(p_j(w))$ with mean $w_i$ and finite variance. Choosing the $w$'s to be close to one, we made the branching process near-critical. In §5.2 we calculated the principal eigenvalue of the mean matrix of this branching process. Using the principal eigenvalue and the corresponding eigenvector, we wrote down the generator of the Feller diffusion limit of the rescaled process. As a result, we were able to describe the evolution of allele frequencies in a population evolving according to this branching process model. Each individual in the population carries one of the two different alleles ($A$ or $a$). This allele is then transmitted to her offspring. Assuming that mutation is not present, we obtained the following theorem.

**Theorem 7.2.** *The proportion of A-type alleles converges weakly to the diffusion $q_t$ given by*

$$dq_t = \sigma\sqrt{\frac{q_t(1-q_t)}{s_t}}\,dB_t$$
$$ds_t = \beta s_t\,dt + \sigma\sqrt{s_t}\,dB_t,$$

*and stopped if $q_t$ hits 0 or 1, in other words, whenever the allele A fixes or dies*

*out. Here $s_t$ corresponds to the total size of the population, while $B_t$ is a standard Brownian motion. The constant $\beta$ depends on both the weight sequence $w = (1 + \epsilon_j/N)$ and the the offspring number distribution $(p_j(1)$:*

$$\beta = \sum_{i=1}^{\infty} i\epsilon_i p_i(1) \text{ and}$$
$$\sigma^2 = \sum_{i=1}^{\infty} i(i-1)p_i(1).$$

This result corresponds to the familiar case where the population evolves according to a single-type near-critical Galton-Watson process with offspring number mean $1 + \beta/N$ and variance $\sigma^2$. Our contribution here was the calculation of the explicit value of $\beta$ in the multi-type context of cultural transmission of fitness.

In Part IV we described a subcritical branching diffusion whose underlying motion was chosen in such a way that we could interpret the particles' spatial position as their expected family size. In this model for CTF, greater family size is associated with strictly greater reproductive fitness, reflecting the assumption that a mother's benefit from cooperation with her siblings strictly increases with their numbers without bound. The spatial motion has a negative drift term modelling the assumption that the fertility of an ageing mother slowly decreases in average over time, or that the number of elder siblings has a detrimental effect on the number of an individual's children. We saw that, in this case, the genealogies qualitatively differ from the ones in the first two models. In §6.2 we applied Theorem 3 in Berestycki et al. (2013) to deduce that the model's limit dual is a time-scaled $U$-coalescent: multiple mergers of the lineages occur. This demonstrated that the notion of CTF can lead out from the familiar framework not only in case of a constant size population (look back to Part II) but also branching process models, as Austerlitz and Heyer (1998)

originally predicted.

One critique of the above branching diffusion is that, because of the negative drift in the underlying spatial motion, late born offspring do not inherit their mother's family size. A better alternative would be a spatial motion $X_t$ with zero drift and the coefficient $\sigma(X_t)$ of the $dB_t$ term set so that spatial position corresponds to family size, if individuals are killed at an exponential rate $\rho(X_t)$ that also depends on their position. Calculating $\rho$ and $\sigma$ is relatively easy, applying the Feyman-Kac formula. Finding the associated coalescent process is much harder.

# References

F. Austerlitz and E. Heyer. Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. *Proceedings of the National Academy of Sciences*, 95(25):15140, 1998.

A.G. Bell. Is race suicide possible? *Journal of Heredity*, 11(8):339–341, 1920.

E.M.S. Belle, U. Ramakrishnan, J.L. Mountain, and G. Barbujani. Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proceedings of the National Academy of Sciences*, 103(21):8012–8017, 2006.

J. Berent. Relationship between family sizes of two successive generations. *The Milbank Memorial Fund Quarterly*, 31(1):39–50, 1953.

J. Berestycki, N. Berestycki, and J. Schweinsberg. The genealogy of branching Brownian motion with absorption. *The Annals of Probability*, 41(2):527–618, March 2013.

M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger. Alpha-stable branching and Beta-coalescents. *Electronic Journal of Probability*, 10(9), 2005.

M.G.B. Blum, E. Heyer, O. François, and F. Austerlitz. Matrilineal fertility inheritance detected in hunter–gatherer populations using the imbalance of gene genealogies. *PLoS Genetics*, 2(8):122, 2006.

A. Budhiraja and D. Reinhold. Some asymptotic results for near critical branching processes. *Communications on Stochastic Analysis*, 4(1):91–113, 2010.

P. Buri. Gene frequency in small populations of mutant Drosophila. *Evolution*, pages 367–402, 1956.

R.B. Campbell. The coalescent time in the presence of background fertility selection. *Theoretical Population Biology*, 55(3):260, 1999.

R. Chaix, F. Austerlitz, T. Hegay, L. Quintana-Murci, and E. Heyer. Genetic traces of east-to-west human expansion waves in Eurasia. *American Journal of Physical Anthropology*, 136(3):309–317, 2008.

R. Durrett. *Stochastic Calculus: a Practical Introduction*. CRC, 1996.

A. Etheridge. *Some Mathematical Models from Population Genetics: École D'éte de Probabilités De Saint-flour XXXIX-2009*, volume 2012. Springer-Verlag Berlin Heidelberg, 2011.

S.N. Ethier and T.G. Kurtz. *Markov processes: Characterization and convergence*. 1986.

W.J. Ewens. *Mathematical Population Genetics: 1. Theoretical Introduction, Second Edition*, volume 27. Springer-Verlag New York, 2004.

R.A. Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, 1930.

Y.-X. Fu. Exact coalescent for the Wright-Fisher model. *Theoretical Population Biology*, 69(4):385–394, 2006.

J.H. Gillespie. *Population Genetics, A Concise Course, Second Edition*. The Johns Hopkins University Press, 2004.

A. Helgason, B. Hrafnkelsson, J.R. Gulcher, R. Ward, and K. Stefánsson. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies:

evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *The American Journal of Human Genetics*, 72(6):1370–1388, 2003.

E. Heyer, A. Sibert, and F. Austerlitz. Cultural transmission of fitness: genes take the fast lane. *Trends in Genetics*, 21(4):234–239, 2005.

O. Hössjer. Coalescence theory for a general class of structured populations with fast migration. *Advances in Applied Probability*, 43(4):1027–1047, 2011.

J.A. Hostetler. *Hutterite Society*. The John Hopkins University Press, Baltimore, 1997.

R.R. Huestis and A. Maxwell. Does family size run in families? *Journal of Heredity*, 23(2):77, 1932.

A. Joffe and M. Metivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, 18:20–65, 1986.

M. Kimmel and D.E. Axelrod. *Branching Processes in Biology*. Springer, Berlin, 2002.

E. Kreyszig. *Introductory Functional Analysis with Applications, Revised Edition*. Wiley Classics Library, 1989.

A. Lambert. Population Dynamics and Random Genealogies. *Stochastic Models*, 24 (01):45–163, 2008.

T.R. Malthus. *An Essay on the Principle of Population as it Affects the Future Improvement of Society, with Remarks on the Speculations of Mr. Godwin, Mr. Condorcet, and Other Writers*. London, J. Johnson, in St. Paul's Church-Yard, 1798.

M. Möhle. A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Advances in Applied Probability*, 30:493–512, 1998.

M. Murphy. Is the relationship between fertility of parents and children really weak? *Social Biology*, 1-2(46):122–145, 2007.

T. Nagylaki. The strong-migration limit in geographically structured populations. *Journal of Mathematical Biology*, 9(2):101–114, 1980.

M. Nei and M. Murata. Effective population size when fertility is inherited. *Genetical Research*, (8):257–260, 1966.

M. Nordborg and S.M. Krone. *Separation of time scales and convergence to the coalescent in structured populations, Modern Developments in Theoretical Population Genetics, The Legacy of Gustave Malécot*. Oxford University Press, 2002.

B. Øksendal. *Stochastic Differential Equations*. Springer, Berlin, 6th edition, 2003.

K. Pearson and A. Lee. On the inheritance of fertility in mankind. *Royal Society of London Philosophical Transactions*, 192(Series A):941, 1899.

E.B. Poulton. Notes upon, or suggested by, the colours, markings and protective attitudes of certain lepidopterous larvae and pupae, and of a phytophagous hymenopterous larva. *Transactions of the Entomological Society of London*, (1884): 27–60, 1884.

L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales, Volume 1 Foundations*. Cambridge University Press, 2000.

C.R. Scriver. Human genetics: lessons from Quebec populations 1. *Annual Review of Genomics and Human Genetics*, 2(1):69–101, 2001.

A. Sibert, F. Austerlitz, and E. Heyer. Wright-Fisher revisited: the case of fertility correlation. *Theoretical Population Biology*, 62(2):181–197, 2002.

M. Tremblay and H. Vézina. Genealogical analysis of maternal and paternal lineages in the Quebec population. *Human Biology*, 82(2):179–198, 2010.

S. Wright. Evolution in Mendelian populations. *Genetics*, 16(2):97, 1931.

# A Sufficient conditions for existence of a time-stationary family size distribution in Part I

*Proof of Theorem 3.5.* We apply Banach's Fixed Point Theorem with the function $f : [0,1]^{K-1} \to [0,1]^{K-1}$, defined in §3.2. Let $x, y \in [0,1]^{K-1}$, then we have

$$\begin{aligned}
\|f(x) - f(y)\|_\infty &= \max_{1\le j\le K-1} \left| \sum_{m=1}^{K^*} x_m \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - y_m \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right| \\
&= \max_{1\le j\le K-1} \left| \sum_{m=1}^{K^*} x_m \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - y_m \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} + \right. \\
&\qquad \left. + y_m \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - y_m \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right|; \\
&\text{as } x_{K^*} = 1 - \sum_{m=1}^{K-1} x_m \\
&= \max_{1\le j\le K-1} \left| \sum_{m=1}^{K-1} (x_m - y_m) \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} \right. \\
&\qquad - \sum_{m=1}^{K-1} (x_m - y_m) \frac{r_k(x)^j}{(j-1)!} e^{-r_k(x)} + \\
&\qquad \left. + \sum_{m=1}^{K^*} y_m \left( \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right) \right|; \\
&\text{by the triangle inquality} \\
&\le \max_{1\le j\le K-1} \left( \max_{1\le m\le K-1} |x_m - y_m| \sum_{m=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_k(x)^j}{(j-1)!} e^{-r_k(x)} \right| + \right. \\
&\qquad \left. + \max_{1\le m\le K} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right| \sum_{m=1}^{K*} y_m \right); \\
&\text{as } \sum_{m=1}^{K*} y_m = 1 \\
&\le \|x - y\|_\infty \max_{1\le j\le K-1} \sum_{m=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_K(x)^j}{(j-1)!} e^{-r_K(x)} \right| +
\end{aligned}$$

$$+ \max_{\substack{1\le j\le K-1\\ 1\le m\le K}} \left| \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!}e^{-r_m(y)} \right|$$

$$=: s\|x-y\|_\infty \text{for some } s \in \mathbf{R}^+.$$

The mapping $f$ is a contraction on $[0,1]^{K-1}$ if $s < 1$. Enforcing this condition on the inequality above completes the proof of the first statement. Manipulating $\|f(x) - f(y)\|_1$ in the same way, leads to the second sufficient condition.

$$\begin{aligned}
\|f(x)-f(y)\|_1 &= \sum_{j=1}^{K-1} \left| \sum_{m=1}^{K^*} x_m \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} - y_m \frac{r_m(y)^j}{(j-1)!}e^{-r_m(y)} \right| \\
&= \sum_{j=1}^{K-1} \left| \sum_{m=1}^{K^*} x_m \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} - y_m \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} + \right. \\
&\qquad \left. + y_m \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} - y_m \frac{r_m(y)^j}{(j-1)!}e^{-r_m(y)} \right|; \\
&\text{as } x_{K^*} = 1 - \sum_{m=1}^{K-1} x_m \text{ and } y_{K^*} = 1 - \sum_{m=1}^{K-1} y_m \\
&= \sum_{j=1}^{K-1} \left| \sum_{m=1}^{K-1} (x_m - y_m) \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} \right. \\
&\qquad - \frac{r_K(x)^j}{(j-1)!}e^{-r_K(x)} \sum_{m=1}^{K-1}(x_m - y_m) + \\
&\qquad \left. + \sum_{m=1}^{K^*} y_m \left( \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!}e^{-r_m(y)} \right) \right|; \\
&\text{by the triangle inequality and as the } y_m\text{'s are non-negative} \\
&\le \sum_{j=1}^{K-1} \sum_{m=1}^{K-1} |x_m - y_m| \left| \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} - \frac{r_K(x)^j}{(j-1)!}e^{-r_K(x)} \right| + \\
&\qquad + \sum_{j=1}^{K-1} \max_{1\le m\le K} \left| \frac{r_m(x)^j}{(j-1)!}e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!}e^{-r_m(y)} \right| \sum_{m=1}^{K^*} y_m; \\
&\text{as } \sum_{m=1}^{K^*} y_m = 1
\end{aligned}$$

$$= \sum_{m=1}^{K-1} |x_m - y_m| \sum_{j=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_K(x)^j}{(j-1)!} e^{-r_K(x)} \right| +$$

$$+ \max_{1 \le m \le K} \sum_{j=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right|$$

$$\le \|x - y\|_1 \max_{1 \le m \le K} \sum_{j=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_K(x)^j}{(j-1)!} e^{-r_K(x)} \right| +$$

$$+ \max_{1 \le m \le K} \sum_{j=1}^{K-1} \left| \frac{r_m(x)^j}{(j-1)!} e^{-r_m(x)} - \frac{r_m(y)^j}{(j-1)!} e^{-r_m(y)} \right|$$

$=: s\|x - y\|_1$ for some $s \in [0, 1)$. □

# B Parameters for the approximations in Part II

| Parameter | $\beta$ | $\gamma$ | $\delta$ |
|---|---|---|---|
| a | 0.25 | 0.435 | -1.4 |
| b | 3.1 | 2.75 | 2.75 |
| c | 1 | 2 | 9 |
| d | 1.3 | 0 | 0 |

**Parameters of the functions $\beta$, $\gamma$ and $\delta$ from §4.1**

| Parameter $n = 255$ | $n = \lfloor N/50 \rfloor$ | $n = \lfloor \sqrt{N} \rfloor$ | $n = N$ |
|---|---|---|---|
| $a$ | 0.1313 | 0.1896 | 0.2953 |
| $b$ | 7.7281 | 7.9306 | 8.8120 |
| $c$ | 1.1172 | 1.3549 | 1.4340 |

**Parameters of the approximation of $\mathrm{E}[I']$ with different sample sizes $n$ from §4.3**

**Average number of families by their size** $\lim_{\alpha\to\infty} \mathrm{E}[f_t(k)]$ at $k = 1, 2, \ldots 20$.

| | |
|---|---|
| 90.7216 | 4.0033 |
| 50.1512 | 3.0224 |
| 28.4944 | 2.1893 |
| 18.8468 | 1.5168 |

| | |
|---|---|
| 13.9967 | 1.0031 |
| 11.1554 | 0.6323 |
| 9.2043 | 0.3797 |
| 7.6647 | 0.2171 |
| 6.3264 | 0.1181 |
| 5.1102 | 0.0612 |

# C Matlab code for Part II

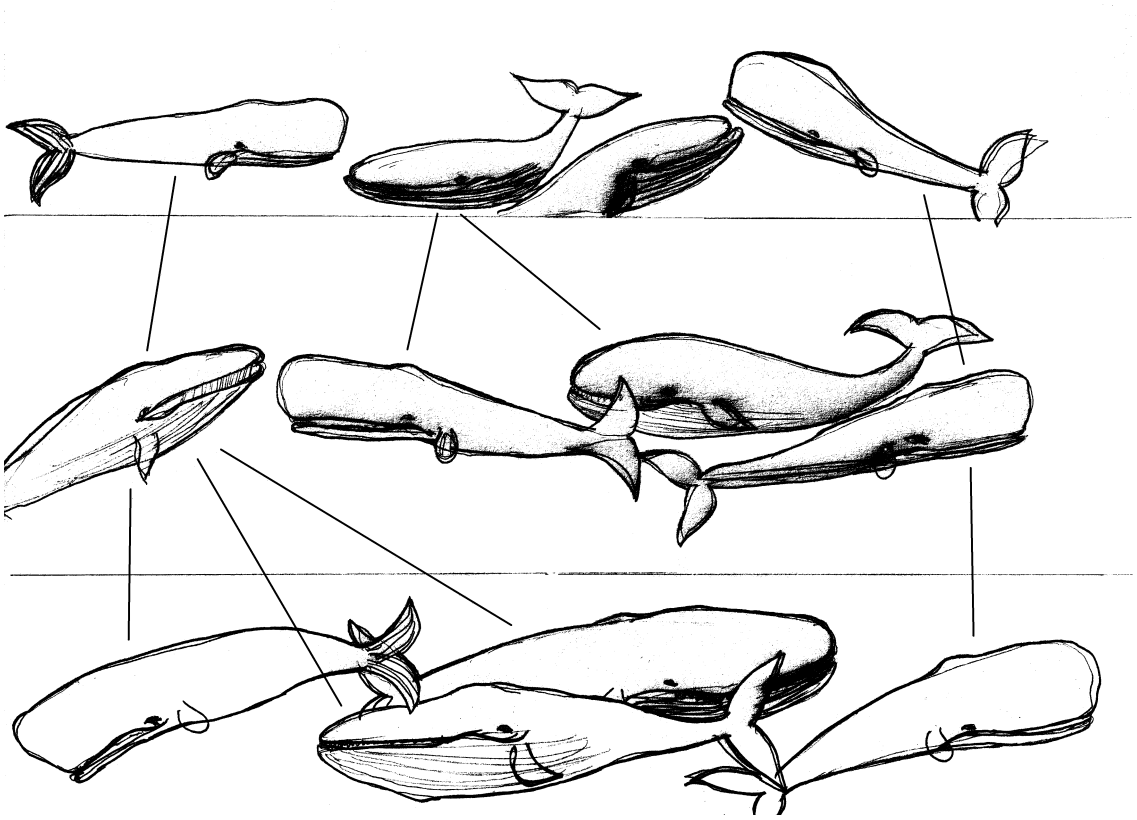


Figure 17: Mother-daughter relationships in a whale population. The individual whales are classified by their family size. For example, the second generation consists of two singletons and two whales of family size two.