

STOCHASTIC MODELS FOR GENETIC EVOLUTION

Frank den Hollander
Mathematical Institute
Leiden University
P.O. Box 9512
2300 RA Leiden
The Netherlands
denhollander@math.leidenuniv.nl

December 2005

PREFACE

The goal of this course is to present a series of elementary stochastic models from population dynamics capable of describing rudimentary aspects of DNA sequence evolution. Most of the course focusses on the Wright-Fisher model and its variations: one colony with resampling, mutation and/or selection, and many colonies with migration.

Key words: Wright-Fisher model, Moran-model, coalescent, duality, space-time scaling, diffusion, resampling, mutation, selection, migration, renormalization, universality.

The course uses as input material various parts of the monograph:

- (1) R. Durrett, *Probability Models for DNA Sequence Evolution*, Springer, New York, 2002.

Further information can be obtained from the monographs:

- (2) W. Ewens, *Mathematical Population Genetics* (2nd. ed.), Springer, Berlin, 2005.
- (3) J. Hein, M.H. Schierup and C. Wiuf, *Gene Genealogies, Variation and Evolution*, Oxford University Press, Oxford, 2005.
- (4) S.H. Rice, *Evolutionary Theory: Mathematical and Conceptual Foundations*, Sinauer, Sunderland MA, 2004.

DISCLAIMER: The present notes are preliminary and need to be further developed. Chapter 1 (genetic background) needs to be expanded, while in Chapters 2 and 3 (Wright-Fisher and its variations) applications of the theory need to be worked out.

Contents

1	Genetic background	4
2	Evolution of a Wright-Fisher population	6
2.1	Standard Wright-Fisher model	6
2.1.1	WF-model	6
2.1.2	WF-diffusion	9
2.1.3	Dual to WF-diffusion	10
2.1.4	Moran-model	11
2.2	Wright-Fisher with mutation	12
2.2.1	Mutation	12
2.2.2	Weak mutation	15
2.3	Wright-Fisher with mutation and infinitely many types	17
2.3.1	Infinite alleles model	17
2.3.2	Ewens sampling formula	19
2.3.3	GEM-distribution	20
2.4	Wright-Fisher with mutation and infinitely many sites	22
2.4.1	Infinite sites model	22
2.4.2	Segregating sites	24
2.5	Wright-Fisher with selection	26
2.5.1	Selection	26
2.5.2	Weak selection	28
3	Interacting Wright-Fisher populations	30
3.1	The stepping stone model	30
3.1.1	Lineages	31
3.1.2	Regime I	33
3.1.3	Regime II	33
3.2	Hierarchical models	38
3.2.1	Hierarchically interacting diffusions	38
3.2.2	Hierarchy of space-time scales	39
3.2.3	Local mean-field limit	40
3.2.4	Renormalization transformation	41

1 Genetic background

The hereditary information of most living organisms is carried by *DNA molecules*. DNA usually consists of two complementary chains twisted around each other to form a double helix. Each chain is a linear sequence of four *nucleotides*:

A = adenine
C = cytosine
G = guanine
T = thymine

which pair in the combinations $A - T$ and $C - G$ via hydrogen bonds. For instance, the yeast genome consists of a sequence of 1.2×10^7 nucleotides that occur with frequencies

A: 0.3090
C: 0.1917
G: 0.1913
T: 0.3078

0.9998

Note here that the pairs $A - T$ and $C - G$ do not occur with equal frequency.

DNA plays the role of a *genetic codebook*. Embedded in the long string of nucleotides that make up the *genome* there are so-called *protein-coding genes*. These genes are “transcribed” into (messenger) RNA, which subsequently is “translated” into proteins. (RNA uses uracil U instead of thymine T and is single-stranded.)

Amino acids are the basic structural units of *proteins*. All proteins in all organisms, from bacteria to humans, are constructed from 20 different amino acids. Each of these is coded by triplets of adjacent nucleotides, called *codons*. Most amino acids are coded by more than one triplet. There are 3 triplets that are called *stop codons*, because they are not codes for amino acids but for terminating transcription. The remaining $4^3 - 3 = 61$ triplets are all codes for one of the 20 amino acids.

DNA is situated in the *chromosomes*, which reside in the *nucleus* of the cells. Much of the information encoded in DNA apparently serves no purpose. The human genome has 3×10^9 nucleotides, in which are embedded only 6×10^4 protein-coding genes. Each such gene consist of a large number of nucleotides.

DNA is subject to different types of evolution. For instance, nucleotides may be substituted by others: $A \leftrightarrow G$ and $C \leftrightarrow T$ are called *transitions*, the other substitutions are called *transversions*. It turns out that transitions occur 10-20 times more frequently than transversions. The reason for this is that only few transitions affect the coding role of the triplet, while roughly half of the transversions do. Thus, the coding roles are relatively stable.

Lower organisms (such as bacteria) are *haploid*, i.e., the chromosomes carry only one copy of the genetic material. Most higher organisms (such as humans) are *diploid*, i.e., the chromosomes carry two copies of the genetic material. Some plants have more than two copies. When haploid individuals reproduce, there is *one parent* that passes copies

of its genetic material to its offspring. When diploid individuals reproduce, there are *two parents*, each of which contributes one of each of its pairs of chromosomes. Before being passed on, the two copies of a chromosome may exchange genetic material, a phenomenon called *recombination*.

In this course we will focus on the so-called Wright-Fisher model for the evolution of a population of genes and a number of its variations. In Chapter 2 we consider the single-colony Wright-Fisher model, describing the notions of *resampling*, *mutation* and *selection*. In Chapter 3 we look at the multiple-colony Wright-Fisher model, describing the notion of *migration*.

2 Evolution of a Wright-Fisher population

In this chapter we introduce the Wright-Fisher model and look at a number of its variations. The simplest version of the WF-model dates back to the 1940's and is used to describe the evolution of a population of genes (referred to as “individuals”) of *two types*, called A and a . These types are *neutral*, i.e., their reproductive success does not depend on the type, and their reproduction is *random*. In Section 2.1 we describe the model and study a number of its key properties. In Section 2.2 we add mutation. In Sections 2.3–2.4 we look at the infinite alleles and the infinite sites version of WF, while in Section 2.5 we add selection.

2.1 Standard Wright-Fisher model

2.1.1 WF-model

Consider N diploid individuals, each carrying 2 copies of a specific *genetic locus* (a location of interest in the genome). We think of these as $2N$ haploid individuals, each carrying 1 copy of the locus. Each individual can be of two *types*, A and a , called *alleles* (two different pieces of genetic information at the locus).

Suppose that at each time unit each individual *randomly chooses another individual from the population and adopts its type* (“parallel updating”). This is called *resampling*, and is a form of random reproduction due to the occurrence of nucleotide substitutions. Suppose that all individuals update independently from each other and independently of how they updated at previous times. We are interested in the evolution of the following quantity:

$$X_n = \text{number of } A\text{'s at time } n. \quad (2.1.1)$$

Note that the total population size, $2N$, is fixed. In what follows we will be mainly interested in the case $N \gg 1$. Later we will even pass to the limit $N \rightarrow \infty$.



Fig. 1. Ω , the state space of the Wright-Fisher model.

The sequence $X = (X_n)_{n \in \mathbb{N}_0}$ (with $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$) is a Markov chain on state space $\Omega = \{0, 1, \dots, 2N\}$ (see Fig. 1) with transition kernel

$$p(i, j) = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(\frac{2N-i}{2N}\right)^{2N-j}, \quad i, j \in \Omega. \quad (2.1.2)$$

Indeed, given that at time n the number of individuals of type A equals i , in order to get j individuals of type A at time $n + 1$, exactly j individuals have to choose an

individual of type A and $2N - j$ individuals have to choose an individual of type a . The latter occur with probabilities given by the second and third factor. The number of ways to choose j individuals from $2N$ is given by the first factor. As initial condition we may pick any $X_0 \in \Omega$, say $X_0 = N$.

The states 0 and $2N$ are *traps*: $p(0,0) = p(2N,2N) = 1$. Since all other states communicate, the process eventually gets trapped, in which case all individuals have the same type (all A or all a). Thus, *eventually genetic variability is lost through chance*. We are interested in computing the time until fixation,

$$\tau = \min\{n \in \mathbb{N}_0 : X_n = 0 \text{ or } X_n = 2N\}, \quad (2.1.3)$$

the probability of fixation at 0 ,

$$\mathbb{P}(X_\tau = 0), \quad (2.1.4)$$

and we want to understand how the process behaves prior to τ . The answer will depend on N and X_0 .

Lemma 2.1.1 $\mathbb{P}(X_\tau = 2N \mid X_0 = i) = \frac{i}{2N}$.

Proof. Abbreviate $p_i = \frac{i}{2N}$. Then $p(i, \cdot) = \text{BIN}(2N, p_i)(\cdot)$, the binomial distribution with $2N$ trials and success probability p_i . Hence

$$\mathbb{E}(X_{n+1} \mid X_n = i) = 2Np_i = i = X_n. \quad (2.1.5)$$

What this says is that our Markov chain X is a *martingale* (= a fair game). Since the state space Ω is finite, standard martingale theory tells us that X converges to a limit, i.e.,

$$X_n \longrightarrow X_\infty \quad \text{a.s. as } n \rightarrow \infty \quad (2.1.6)$$

with X_∞ some random variable. In our case, obviously $X_\infty = X_\tau$. Next, we write

$$i = \mathbb{E}(X_n \mid X_0 = i) = \mathbb{E}(X_\tau 1\{\tau \leq n\} \mid X_0 = i) + \mathbb{E}(X_n 1\{\tau > n\} \mid X_0 = i), \quad (2.1.7)$$

where we use that $X_n = X_\tau$ when $\tau \leq n$. Now let $n \rightarrow \infty$, and use that $\mathbb{P}(\tau < \infty) = 1$ and Ω is finite, to see that the first term tends to $\mathbb{E}(X_\tau \mid X_0 = i)$ and the second term tends to zero. Thus, we have

$$\mathbb{E}(X_\tau \mid X_0 = i) = i. \quad (2.1.8)$$

But

$$\mathbb{E}(X_\tau \mid X_0 = i) = 0 \times \mathbb{P}(X_\tau = 0 \mid X_0 = i) + 2N \times \mathbb{P}(X_\tau = 2N \mid X_0 = i). \quad (2.1.9)$$

Combine (2.1.8–2.1.9) to get the claim. ■

To investigate $\mathbb{E}(\tau)$, we consider the quantity

$$H_n = \frac{2X_n(2N - X_n)}{2N(2N - 1)}. \quad (2.1.10)$$

This is called the *heterozygosity* of the population at time n and equals the probability that two individuals randomly drawn from the population at time n are of different type.

Lemma 2.1.2 $\mathbb{E}(H_n) = (1 - \frac{1}{2N})^n H_0$.

Proof. Randomly draw two individuals at time n . Draw the two *backward ancestral paths* of these two individuals, i.e., the paths labelling the individuals that were successively chosen as ancestors at all previous times (see Fig. 2).

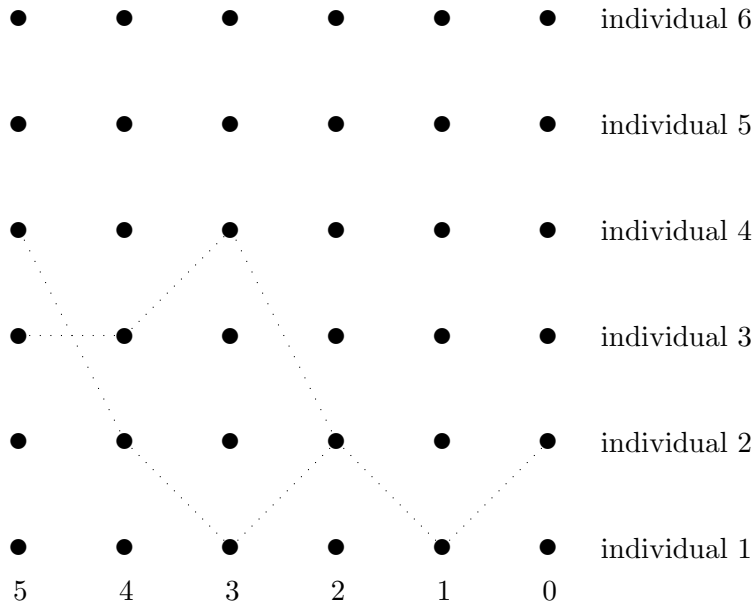


Fig. 2. Two backward ancestral paths for $N = 3$ and $n = 5$. Time runs horizontally to the left.

These paths, which are random, contain the full *genealogical history* of the two individuals. When read backwards in time, these paths behave as two *coalescing random walks* on $\{1, \dots, 2N\}$, the labelling space of the population: they jump randomly and merge into one upon meeting. At each unit of time, the two random walks have probability $1 - \frac{1}{2N}$ to meet, since they jump uniformly between the labels. The probability that they do not coalesce up to time n , which is time 0 in the forward sense, equals $(1 - \frac{1}{2N})^n$. Clearly, the two individuals at time n are of different type if and only if their ancestral lineages do not coalesce and at time 0 are of different type. The latter has probability H_0 , the heterozygosity of the initial state X_0 . ■

Since

$$\tau = \min\{n \in \mathbb{N}_0 : H_n = 0\}, \quad (2.1.11)$$

we see from Lemma 2.1.2 that $\mathbb{E}(\tau)$ is of order N .

2.1.2 WF-diffusion

Lemma 2.1.2 suggests that it is of interest to consider the following *space-time rescaling* of our process:

$$Y_t^{(N)} = \frac{1}{2N} X_{[2Nt]}^{(N)}, \quad t \geq 0. \quad (2.1.12)$$

Here, we add an upper index (N) to exhibit the underlying N -dependence, and space is shrunk by a factor $2N$ while time is blown up by a factor $2N$. We expect that, in the limit as $N \rightarrow \infty$, if the initial condition scales properly,

$$Y_0^{(N)} \Longrightarrow Y_0 \quad \text{as } N \rightarrow \infty, \quad (2.1.13)$$

then the whole process scales properly,

$$(Y_t^{(N)})_{t \geq 0} \Longrightarrow (Y_t)_{t \geq 0} \quad \text{as } N \rightarrow \infty, \quad (2.1.14)$$

where \Longrightarrow denotes convergence in distribution. In other words, we expect the rescaled process to converge to a limiting process, living on state space $[0, 1]$ and evolving in continuous time. This limiting process must be a Markov process.

Theorem 2.1.3 *The scaling in (2.1.14) subject to (2.1.13) holds true, with $(Y_t)_{t \geq 0}$ the diffusion process on $[0, 1]$ given by the stochastic differential equation*

$$dY_t = \sqrt{Y_t(1 - Y_t)} dW_t, \quad (2.1.15)$$

where $(W_t)_{t \geq 0}$ is standard Brownian motion (the diffusion process with generator $\frac{1}{2} \frac{d^2}{dy^2}$).

Proof. An informal proof runs as follows. Both $(Y_t^{(N)})_{t \geq 0}$ and $(W_t)_{t \geq 0}$ are Markov processes. Compute, for $y \in (0, 1)$ and small $\Delta t > 0$,

$$\begin{aligned} \mathbb{E} \left(Y_{t+\Delta t}^{(N)} - Y_t^{(N)} \mid Y_t^{(N)} = y \right) &= 0, \\ \mathbb{E} \left(\left[Y_{t+\Delta t}^{(N)} - Y_t^{(N)} \right]^2 \mid Y_t^{(N)} = y \right) &= y(1 - y)\Delta t + o(\Delta t), \end{aligned} \quad (2.1.16)$$

which arise from the observation that $\text{BIN}(2N, p_i)$ has mean $2Np_i$ and variance $2Np_i(1 - p_i)$. Since $(W_t)_{t \geq 0}$ satisfies the same relations without the factor $y(1 - y)$, the claim follows from standard diffusion theory. A formal proof requires more advanced stochastic analysis (see e.g. Ethier and Kurtz [5]) ■

The limiting process defined by (2.1.15) is called the *Wright-Fisher diffusion*. Think of this as a standard Brownian motion running at a “local speed” given by a *diffusion function* $g: [0, 1] \rightarrow [0, \infty)$, in our case $g(y) = y(1 - y)$ (see Fig. 3). Note that 0 and 1 are traps: the WF-diffusion stops when it reaches the boundary of $[0, 1]$. Equation (2.1.15) has a unique strong solution, i.e., there is a unique path $t \mapsto Y_t$ that is measurable w.r.t. the canonical filtration associated with the Brownian motion.

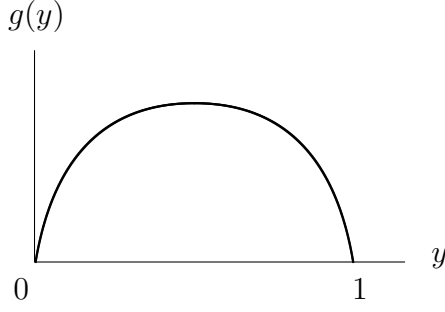


Fig. 3. Wright-Fisher diffusion function.

The fact that Brownian motion has generator $\frac{1}{2} \frac{d^2}{dy^2}$ amounts to the assertion that

$$\mathbb{E}(f(W_t)) - \mathbb{E}(f(W_0)) = \int_0^t \frac{1}{2} f''(W_s) ds \quad (2.1.17)$$

for all f that are twice continuously differentiable. The WF-diffusion has generator $y(1-y) \frac{1}{2} \frac{d^2}{dy^2}$, i.e.,

$$\mathbb{E}(f(Y_t)) - \mathbb{E}(f(Y_0)) = \int_0^t Y_s(1-Y_s) \frac{1}{2} f''(Y_s) ds. \quad (2.1.18)$$

2.1.3 Dual to WF-diffusion

The WF-diffusion describes the WF-model on large space-time scales. Even though there is no easy explicit formula for Y_t in terms of W_s , $0 \leq s \leq t$, (2.1.15) has the advantage of being easier to manipulate in computations than the original Markov chain. To illustrate this advantage, we next turn to the notion of *duality*.

Theorem 2.1.4 *Let $(N_t)_{t \geq 0}$ be the death process on $\mathbb{N} = \{1, 2, \dots\}$ where transitions from n to $n-1$ occur at rate $\frac{1}{2}n(n-1)$. Then*

$$\mathbb{E}([Y_t]^n \mid Y_0 = y) = \mathbb{E}(y^{N_t} \mid N_0 = n) \quad \forall y \in [0, 1], n \in \mathbb{N}, t \geq 0. \quad (2.1.19)$$

Proof. Abbreviate

$$\begin{aligned} a_t(y, n) &= \mathbb{E}([Y_t]^n \mid Y_0 = y), \\ b_t(y, n) &= \mathbb{E}(y^{N_t} \mid N_0 = n). \end{aligned} \quad (2.1.20)$$

Let $f(x) = x^n$. Then it follows from (2.1.15) that

$$\begin{aligned} a_t(y, n) - a_0(y, n) &= \mathbb{E}(f(Y_t) - f(Y_0) \mid Y_0 = y) \\ &= \mathbb{E} \left(\int_0^t Y_s(1-Y_s) \frac{1}{2} f''(Y_s) ds \mid Y_0 = y \right) \\ &= \mathbb{E} \left(\int_0^t Y_s(1-Y_s) \frac{1}{2} n(n-1) [Y_s]^{n-2} ds \mid Y_0 = y \right) \\ &= \int_0^t \binom{n}{2} [a_s(y, n-1) - a_s(y, n)] ds, \end{aligned} \quad (2.1.21)$$

where the second equality uses Itô's formula from stochastic analysis. Equation (2.1.21) is an integral recursion relation for $a_t(y, n)$. However, after differentiating the right-hand side of (2.1.21) with respect to t , we see from the definition of $(N_t)_{t \geq 0}$ that $b_t(y, n)$ satisfies the same recursion. Since

$$a_0(y, n) = y^n = b_0(y, n) \quad \forall y \in [0, 1], n \in \mathbb{N}, \quad (2.1.22)$$

it follows that

$$a_t(y, n) = b_t(y, n) \quad \forall y \in [0, 1], n \in \mathbb{N}, t \geq 0. \quad (2.1.23)$$

This proves the claim. \blacksquare

What Theorem 2.1.4 says is that the moments of Y_t can be computed from the distribution of N_t , and therefore so can the distribution of Y_t itself. Since the dual process $(N_t)_{t \geq 0}$ is *much simpler* than the WF-diffusion $(Y_t)_{t \geq 0}$, this constitutes a considerable advantage. For instance, we have

$$\mathbb{P}(Y_\infty = 1 \mid Y_0 = y) = \mathbb{E}(Y_\infty \mid Y_0 = y) = \mathbb{E}(y^{N_\infty} \mid N_0 = 1) = y, \quad (2.1.24)$$

where we use (2.1.19) and the fact that $Y_\infty \in \{0, 1\}$ and $N_\infty = 1$. This reflects the result in Lemma 2.1.1. Similarly,

$$\begin{aligned} \mathbb{E}(Y_t(1 - Y_t) \mid Y_0 = y) &= \mathbb{E}(y^{N_t} \mid N_0 = 1) - \mathbb{E}(y^{N_t} \mid N_0 = 2) \\ &= y - [y\mathbb{P}(N_t = 1 \mid N_0 = 2) + y^2\mathbb{P}(N_t = 2 \mid N_0 = 2)] \\ &= y(1 - y)\mathbb{P}(N_t = 2 \mid N_0 = 2) \\ &= y(1 - y)e^{-t}. \end{aligned} \quad (2.1.25)$$

This reflects the result in Lemma 2.1.2.

The dual process $(N_t)_{t \geq 0}$ can in fact be directly traced back to the backward coalescing random walk encountered in the proof of Lemma 2.1.2. Indeed, $(Y_t)_{t \geq 0}$ describes the evolution of the type decomposition of a large population. Suppose that we sample n individuals from this population at time t . The left-hand side of (2.1.19) is the probability to see only type A in the sample. But we can compute this probability differently. If we know that the n individuals in the sample are the descendants of N_t different ancestors at time 0, and that the fraction of type A at time 0 is y , then by averaging over the random genealogy we obtain the right-hand side of (2.1.19) for the probability that all of the N_t ancestors are of type A .

A full description of all the lineages of the n individuals in the sample is called *Kingman's coalescent*. This is a random process taking values in the collection of partitions of $\{1, \dots, n\}$ such that, if the partition has j sets, then at rate $\frac{1}{2}j(j-1)$ two randomly chosen sets in the partition are joined together. See Durrett [4], pp. 10–12.

2.1.4 Moran-model

There is a *continuous-time* version of the WF-model, called the *Moran-model*, in which each individual chooses a random ancestor at rate 1 and adopts its type (“sequential updating”). The resulting process is a birth-death process on state space Ω with transition

rates

$$\begin{aligned} i \rightarrow i + 1 & \text{ at rate } b_i = (2N - i) \frac{i}{2N}, \\ i \rightarrow i - 1 & \text{ at rate } d_i = i \frac{2N - i}{2N}. \end{aligned} \tag{2.1.26}$$

The differences between this model and the discrete-time WF-model (“parallel updating”) are minor. In particular, after space-time scaling the same WF-diffusion occurs as limit.

A straightforward extension of the Moran-model is obtained when the rate at which the individuals choose their ancestor is not 1 but $h(y)$, where y is the state of the system (= the fraction of type A) and $h: [0, 1] \rightarrow [0, \infty)$ is any Lipschitz function. The latter plays the role of an *overall speed up of the rate of resampling*, and models an effect that the population size may have on the evolution rate. Examples are:

$$\begin{aligned} h(y) = 1 & \quad \text{Wright-Fisher model,} \\ h(y) = y(1 - y) & \quad \text{Ohta-Kimura model.} \end{aligned} \tag{2.1.27}$$

The latter example corresponds to an overall rate that is proportional to the heterozygosity of the population (“random selection model”).

Returning to Theorem 2.1.3, we find that the space-time scaling of the h -model is given by:

Theorem 2.1.5 *The h -version of the WF-diffusion satisfies the stochastic differential equation*

$$dY_t = \sqrt{Y_t(1 - Y_t)h(Y_t)} dW_t. \tag{2.1.28}$$

Proof. Elementary. ■

It is expedient to define

$$g(y) = y(1 - y)h(y) \tag{2.1.29}$$

and to write (2.1.28) as

$$dY_t = \sqrt{g(Y_t)} dW_t. \tag{2.1.30}$$

Here, $g: [0, 1] \rightarrow [0, \infty)$ is the *local diffusion function*. In order for (2.1.30) to be properly defined and have a unique strong solution, some restrictions have to be placed on g . We will return to this in Section 3.2.

In general there is no nice dual process for the diffusion given by (2.1.30), unlike for the standard WF-diffusion.

2.2 Wright-Fisher with mutation

2.2.1 Mutation

Suppose that we modify the WF-model in the following manner. At each time unit, each individual, immediately after it has chosen its ancestor and adopted its type, suffers a *type mutation*: type a spontaneously mutates into type A , with probability u , and type

A spontaneously mutates into type a , with probability v (see Fig. 4). Here, $0 < u, v < 1$, and the mutation occurs independently for different individuals.

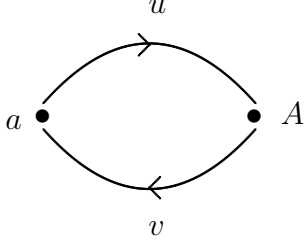


Fig. 4. Two-type mutation.

Our goal is to investigate what effect the mutation has on the behavior of the model. Under the present conditions, $X = (X_n)_{n \in \mathbb{N}_0}$ is a Markov chain on the state space Ω with transition kernel

$$p(i, j) = \binom{2N}{j} (p_i)^j (1 - p_i)^{2N-j}, \quad i, j \in \Omega, \quad (2.2.1)$$

with

$$p_i = \left(\frac{i}{2N} \right) (1 - v) + \left(\frac{2N - i}{2N} \right) u. \quad (2.2.2)$$

Indeed, either an A is drawn (probability $\frac{i}{2N}$) and it does not mutate (probability $1 - v$), or an a is drawn (probability $\frac{2N - i}{2N}$) and it does mutate (probability u). Compare (2.2.1–2.2.2) with (2.1.2).

A first consequence of the presence of mutation is that the traps at 0 and $2N$ disappear: $p(i, j) > 0$ for all $i, j \in \Omega$. Thus, there is no longer loss of genetic variability. In fact, because N is finite, X converges to an *equilibrium* on Ω ,

$$\pi(i) = \mathbb{P}(X_\infty = i), \quad i \in \Omega, \quad (2.2.3)$$

solving the set of equations

$$\sum_{i \in \Omega} \pi(i) p(i, j) = \pi(j), \quad j \in \Omega, \quad (2.2.4)$$

normalized such that $\sum_{i \in \Omega} \pi(i) = 1$. The equilibrium has full support, i.e., $\pi(i) > 0$ for all $i \in \Omega$.

In principle it is possible to compute π , but the formulas are not so pretty. We get some insight by computing moments. We will compute the first two moments, the second only for $N \gg 1$ and $0 < u, v \ll 1$.

Lemma 2.2.1 *The first and second moment of X_∞ are given by*

$$\begin{aligned} \mathbb{E}(X_\infty) &= 2N \left(\frac{u}{u+v} \right), \\ \mathbb{E}(X_\infty(X_\infty - 1)) &\sim 2N(2N - 1) \left\{ \frac{1}{1 + 4N(u+v)} \left(\frac{u}{u+v} \right) + \frac{4N(u+v)}{1 + 4N(u+v)} \left(\frac{u}{u+v} \right)^2 \right\}. \end{aligned} \quad (2.2.5)$$

Proof. Let $\eta_i = 1$ if the i -th individual is A (under the equilibrium distribution) and 0 otherwise. Then

$$X_\infty = \sum_{i=1}^{2N} \eta_i. \quad (2.2.6)$$

By exchangeability, we have

$$\begin{aligned} \mathbb{E}(X_\infty) &= 2N\mathbb{P}(\eta_1 = 1), \\ \mathbb{E}(X_\infty(X_\infty - 1)) &= 2N(2N - 1)\mathbb{P}(\eta_1 = \eta_2 = 1). \end{aligned} \quad (2.2.7)$$

To compute the first probability in (2.2.7), abbreviate $\rho = \mathbb{P}(\eta_1 = 1)$. In equilibrium we have

$$\rho = (1 - v)\rho + u(1 - \rho). \quad (2.2.8)$$

This gives $\rho = \frac{u}{u+v}$ and proves the first line of (2.2.5). To compute the second probability in (2.2.7), note that in equilibrium

$$\mathbb{P}(\eta_1 = \eta_2 = 1) = \chi\rho + (1 - \chi)\rho^2, \quad (2.2.9)$$

where

$$\begin{aligned} \chi &= \text{the probability that two individuals are } \textit{identical by descent}, \\ &\text{i.e., their lineages coalesce before a mutation affect either lineage.} \end{aligned} \quad (2.2.10)$$

Then, for small u, v ,

$$1 - \chi \sim 2\mu + (1 - 2\mu) \left(1 - \frac{1}{2N} \right) (1 - \chi) \quad (2.2.11)$$

with $\mu \sim u + v$ the probability of mutation in one time step. Indeed, if at the first time step backwards there is no mutation in either of the two lineages nor coalescence of the two lineages, then “the game starts all over again”. Hence we find, for small u, v and large N ,

$$1 - \chi = \frac{2\mu}{1 - (1 - 2\mu)(1 - \frac{1}{2N})} \sim \frac{2(u+v)}{2(u+v) + \frac{1}{2N}} = \frac{4N(u+v)}{1 + 4N(u+v)}. \quad (2.2.12)$$

This proves the second line of (2.2.5). ■

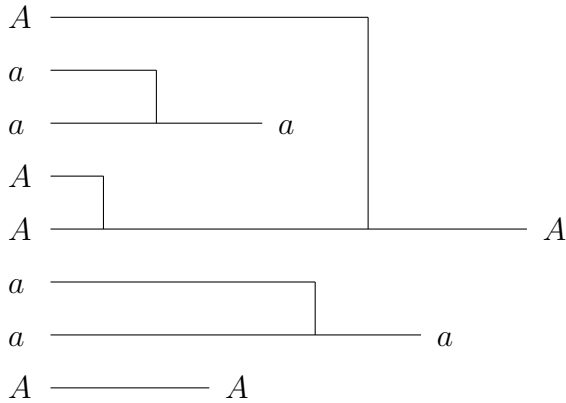


Fig. 5. The coalescent with mutation. Time runs horizontally to the left.

In the presence of mutation, the coalescent (= the backward tree of lineages) modifies in the following manner. A backward path: (i) is “killed and labelled A ” with probability u , (ii) is “killed and labelled a ” with probability v , (iii) jumps to a randomly chosen site with probability $1 - u - v$. Killing a path determines the state of all its descendants (see Fig. 5). Note that if all the backward paths are killed before they reach down to time 0, then the state of the system no longer depends on the initial configuration. This is why the system reaches equilibrium in the presence of mutation.

2.2.2 Weak mutation

Interesting behavior shows up in the limit as $N \rightarrow \infty$, provided we take the mutation probabilities small:

$$u = u(N) = \frac{q}{4N}, \quad v = v(N) = \frac{r}{4N}, \quad q, r > 0. \quad (2.2.13)$$

The reason behind this choice is apparent from Lemma 2.2.1, which shows that $N(u+v)$ is the relevant quantity. Think of q, r as the *mutation rates of the population*. Define, in analogy with (2.1.12),

$$Y^{(N)} = \frac{1}{2N} X_{\infty}^{(N)}. \quad (2.2.14)$$

Then we expect that

$$Y^{(N)} \Longrightarrow Y \quad \text{as } N \rightarrow \infty. \quad (2.2.15)$$

Theorem 2.2.2 *The convergence in (2.2.15) holds true, with Y the random variable on $[0, 1]$ with probability density $\text{BETA}(q, r)$, i.e.,*

$$f(x) = C_{q,r} x^{q-1} (1-x)^{r-1}, \quad x \in [0, 1], \quad (2.2.16)$$

where $C_{q,r} = \Gamma(q+r)/\Gamma(q)\Gamma(r)$ is the normalizing constant (Γ is the Gamma-function).

Proof. The proof proceeds by computing the moments of Y . We will show that

$$\mathbb{E}(Y^k) = \frac{(k+q-1) \times \cdots \times (q+1)q}{(k+q+r-1) \times \cdots \times (q+r+1)(q+r)} = m_k, \quad k \in \mathbb{N}, \quad (2.2.17)$$

which are the moments of $\text{BETA}(q, r)$. The computation of m_k is by induction, namely, we will show that

$$m_k = \frac{k+q-1}{k+q+r-1} m_{k-1}, \quad k \in \mathbb{N}. \quad (2.2.18)$$

Given that $X_{n-1}^{(N)} = i$, the distribution of $X_n^{(N)}$ is $\text{BIN}(2N, p_i)$. Therefore

$$\begin{aligned} \mathbb{E} \left(X_n^{(N)} (X_n^{(N)} - 1) \times \cdots \times (X_n^{(N)} - k + 1) \mid X_{n-1}^{(N)} = i \right) \\ = 2N(2N-1) \times \cdots \times (2N-k+1) (p_i)^k. \end{aligned} \quad (2.2.19)$$

Dividing by $(2N)^k$ and expanding to leading order in $1/2N$, we may rewrite this equation as

$$\begin{aligned} \mathbb{E} \left(\left[\frac{X_n^{(N)}}{2N} \right]^k \mid X_{n-1}^{(N)} = i \right) - \frac{k(k-1)}{4N} \mathbb{E} \left(\left[\frac{X_n^{(N)}}{2N} \right]^{k-1} \mid X_{n-1}^{(N)} = i \right) + O \left(\frac{1}{N^2} \right) \\ = \left[1 - \frac{k(k-1)}{4N} + O \left(\frac{1}{N^2} \right) \right] (p_i)^k, \end{aligned} \quad (2.2.20)$$

where we use that $\sum_{j=0}^{k-1} j = \frac{1}{2}k(k-1)$. Next, expand

$$\begin{aligned} (p_i)^k &= \left(\frac{i}{2N} (1-u-v) + u \right)^k \\ &= \left(\frac{i}{2N} \right)^k + k \left(\frac{i}{2N} \right)^{k-1} \left[\frac{q}{4N} - \left(\frac{i}{2N} \right) \frac{q+r}{4N} + O \left(\frac{1}{N^2} \right) \right], \end{aligned} \quad (2.2.21)$$

where we use (2.2.2) and (2.2.13). Putting $Y_n^{(N)} = X_n^{(N)}/2N$, we may rewrite (2.2.20–2.2.21) as

$$\begin{aligned} \mathbb{E} \left([Y_n^{(N)}]^k \mid Y_{n-1}^{(N)} \right) - \frac{k(k-1)}{4N} \mathbb{E} \left([Y_n^{(N)}]^{k-1} \mid Y_{n-1}^{(N)} \right) + O \left(\frac{1}{N^2} \right) \\ = \left[1 - \frac{k(k-1)}{4N} + O \left(\frac{1}{N^2} \right) \right] \\ \times \left\{ [Y_{n-1}^{(N)}]^k + k [Y_{n-1}^{(N)}]^{k-1} \left[\frac{q}{4N} - [Y_{n-1}^{(N)}] \frac{q+r}{4N} + O \left(\frac{1}{N^2} \right) \right] \right\}. \end{aligned} \quad (2.2.22)$$

Taking the expectation over $Y_{n-1}^{(N)}$, letting $n \rightarrow \infty$ and using that $Y_n^{(N)} \Rightarrow Y^{(N)}$, we obtain the relation

$$\frac{k(k+q+r-1)}{4N} \mathbb{E}([Y^{(N)}]^k) = \frac{k(k+q-1)}{4N} \mathbb{E}([Y^{(N)}]^{k-1}) + O \left(\frac{1}{N^2} \right). \quad (2.2.23)$$

Finally, multiplying by $4N/k$, letting $N \rightarrow \infty$ and using that $Y^{(N)} \implies Y$, we obtain (2.2.18). ■

The qualitative behavior of the probability density in (2.2.16) near the boundaries 0 and 1 is different for $q, r < 1$ and $q, r > 1$ (see Fig. 6).

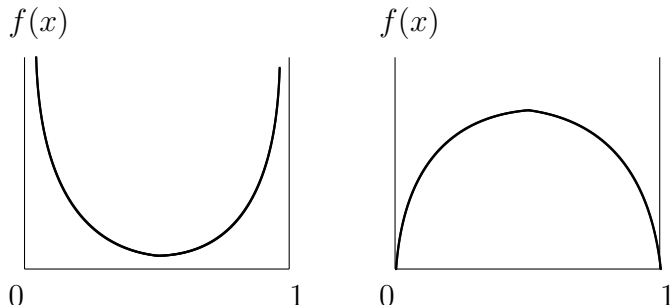


Fig. 6. Qualitative picture of the probability density in (2.2.16) for $q, r < 1$ and $q, r > 1$, respectively.

2.3 Wright-Fisher with mutation and infinitely many types

In this section we consider the modification of the WF-model in which, instead of two types, there are infinitely many types and *each time a mutation occurs it brings in a new type*. The motivation is the following. If a gene consists of 500 nucleotides, then there are $3 \times 500 = 1500$ sequences that can be reached by a single base change. Therefore the probability of returning to the same sequence after two mutations is $1/1500$, which is extremely small. Neglecting this return amounts to considering what is called the “infinite alleles model”.

2.3.1 Infinite alleles model

As before, we consider a population with $2N$ individuals, all starting with, say, type 1. At each unit of time, each individual with probability μ chooses a random ancestor and adopts its type and with probability $1 - \mu$ spontaneously mutates into a new type. All individuals update independently from each other and independently of how they updates at previous times. The first mutation brings in an individual of type 2, the third brings in an individual of type 3, etc. As time proceeds, *new types enter the population and old types die out*. Thus, the types are constantly changing. However, we may expect that after a long time the distribution of the *number* of different types in the population settles down to a limiting distribution. The question is what this distribution is.

We will consider this question in the limit as $N \rightarrow \infty$, with

$$\mu = \mu(N) = \frac{\theta}{4N}, \quad \theta > 0, \quad (2.3.1)$$

where θ may be thought of as the *mutation rate for the population*. In this limit, k lineages

$$\text{coalesce at rate } \frac{1}{2}k(k-1) \text{ and mutate at rate } \frac{1}{2}k\theta. \quad (2.3.2)$$

Draw a random sample of n individuals from the population ($n \ll N \rightarrow \infty$) and ask for

$$K_n = \text{the number of different types in the } n \text{ sample.} \quad (2.3.3)$$

Lemma 2.3.1 *As $n \rightarrow \infty$,*

$$\begin{aligned} \mathbb{E}(K_n) &\sim \theta \log n, \\ \text{Var}(K_n) &\sim \theta \log n. \end{aligned} \quad (2.3.4)$$

Moreover, the central limit applies, i.e., $[K_n - \mathbb{E}(K_n)]/\sqrt{\text{Var}(K_n)}$ converges in distribution to a standard normal random variable.

Proof. The proof uses duality. The dual process is called the *Hoppe urn model*, which is defined as follows:

- An urn contains 1 black ball (of mass θ) and any number of colored balls (of mass 1 each). A ball is selected with a probability *proportional to its mass*. If a *colored* ball is drawn, then an extra ball with the *same color* is put into the urn. If a *black* ball is drawn, then an extra ball with a *new color* is put into the urn.

We start with 1 black ball at time 0. At time n the urn will contain $n + 1$ balls, 1 black and n colored. The state of the urn is the number of different colors and their multiplicity. The key observation is:

- The genealogical relationship between k individuals in the infinite alleles model can be simulated by running the Hoppe urn model for k time steps.

Indeed, the genealogy of the infinite alleles model is described by a *coalescent with killing* similar in spirit to the one encountered after the proof of Lemma 2.2.1 (recall (2.3.2)): lineages perform random walks that coalesce at rate 1 upon meeting, and on each lineage there is an independent Poisson process with rate $\frac{1}{2}\theta$ of mutations (which completely erase the information about the previous type). See the picture on p. 24 of Durrett [4].

With the above duality the proof is easy. Write

$$K_n = \sum_{i=1}^n \eta_i \quad (2.3.5)$$

with $\eta_i = 1$ if the i -th ball added in the Hoppe urn model has a new color and zero otherwise. Clearly, the successive η_i are independent (because at time i the urn always contains $i + 1$ balls) with

$$\mathbb{P}(\eta_i = 1) = \frac{\theta}{\theta + i - 1}. \quad (2.3.6)$$

Consequently,

$$\begin{aligned}\mathbb{E}(K_n) &= \sum_{i=1}^n \mathbb{P}(\eta_i = 1) = \sum_{i=1}^n \frac{\theta}{\theta + i - 1} \\ &\sim \theta \int_{\theta}^{\theta+n} \frac{dx}{x} = \theta[\log(\theta + n) - \log \theta] \sim \theta \log n.\end{aligned}\tag{2.3.7}$$

Similarly,

$$\text{Var}(K_n) = \sum_{i=1}^n \text{Var}(\eta_i) = \sum_{i=1}^n \frac{\theta}{\theta + i - 1} \left(1 - \frac{\theta}{\theta + i - 1}\right) \sim \mathbb{E}(K_n).\tag{2.3.8}$$

This proves (2.3.4). The central limit theorem follows from standard arguments. \blacksquare

Lemma 2.3.1 says that the number of different types in an n sample is rather small, namely, of order $\log n$. Apparently, *new types rapidly eradicate old types*, so that it is hard for different types to coexist in large numbers. It also says that the number of different types is proportional to θ , the population mutation rate. This opens up the possibility to *estimate* θ from *sample data*, which is interesting because typically θ is not known. Lemma 2.3.1 shows that $\hat{\theta}_n = K_n / \log n$ is an asymptotically sharp estimator of θ . Unfortunately, the standard deviation of $\hat{\theta}_n$ decays like $1/\sqrt{\log n}$, which is too slow to get sharp estimates. Still, $\hat{\theta}$ can be used to estimate θ with some confidence interval.

2.3.2 Ewens sampling formula

The next result, which is referred to as *Ewens sampling formula*, describes the full type distribution in the sample. For $i = 1, \dots, n$, let

$$A_i = \text{the number of types that are present precisely } i \text{ times in the sample.}\tag{2.3.9}$$

Note that $\sum_{i=1}^n iA_i = n$.

Theorem 2.3.2 For all $a = (a_i)_{i=1, \dots, n}$ such that $|a| = \sum_{i=1}^n ia_i = n$,

$$\mathbb{P}(A_i = a_i \text{ for } i = 1, \dots, n) = \frac{n!}{\theta(\theta + 1) \times \dots \times (\theta + n - 1)} \prod_{i=1}^n \frac{(\theta/i)^{a_i}}{a_i!}.\tag{2.3.10}$$

Proof. We will make use of the duality with the Hoppe urn model. The proof is by induction on n . Abbreviate $a' = (a'_1, \dots, a'_{n-1})$ and $a = (a_1, \dots, a_n)$. Let $p(a', a)$ denote the transition probability from state a' at time $n - 1$ to state a at time n in the Hoppe urn model. Let $P_\theta(a)$ denote the right-hand side of (2.3.10). We will show that

$$P_\theta(a) = \sum_{a': |a'|=n-1} P_\theta(a')p(a', a) \quad \forall a : |a| = n,\tag{2.3.11}$$

which will imply that $P_\theta(\cdot)$ is the distribution of the Hoppe urn model at time $|\cdot|$.

If $n = 1$, then $a = (a_1) = (1)$ and the claim $P_\theta(1) = 1$ in (2.3.10) is true. Suppose that the state at time $n - 1$ is a and the state at time n is a' . Then there are two possibilities:

- (1) $a_1 = a'_1 + 1$, $a_i = a'_i$ for $i > 1$: a new color is added at time n . For this case, by (2.3.10):

$$\frac{P_\theta(a)}{P_\theta(a')} = \frac{n}{\theta + n - 1} \frac{\theta}{a_1}, \quad p(a', a) = \frac{\theta}{\theta + n - 1}. \quad (2.3.12)$$

- (2) $a_j = a'_j - 1$, $a_{j+1} = a'_{j+1} + 1$, $a_i = a'_i$ for $j \geq 1$ and $i \neq j, j + 1$: a color is added at time n that is present j times at time $n - 1$. For this case, by (2.3.10):

$$\frac{P_\theta(a)}{P_\theta(a')} = \frac{n}{\theta + n - 1} \frac{ja'_j}{(j+1)a_{j+1}}, \quad p(a', a) = \frac{ja'_j}{\theta + n - 1}. \quad (2.3.13)$$

No other transitions are possible. From (2.3.12–2.3.13) we get

$$\begin{aligned} \sum_{a'} \frac{P_\theta(a')}{P_\theta(a)} p(a', a) &= \frac{\theta + n - 1}{n} \frac{a_1}{\theta} \frac{\theta}{\theta + n - 1} + \sum_{j=1}^{n-1} \frac{\theta + n - 1}{n} \frac{(j+1)a_{j+1}}{ja'_j} \frac{ja'_j}{\theta + n - 1} \\ &= \frac{a_1}{n} + \sum_{j=1}^{n-1} \frac{(j+1)a_{j+1}}{n} \\ &= \frac{1}{n} \sum_{j=0}^{n-1} (j+1)a_{j+1} \\ &= 1. \end{aligned} \quad (2.3.14)$$

Hence (2.3.11) indeed holds true, which advances the induction and shows that (2.3.10) holds for all n . \blacksquare

By rewriting (2.3.10) as

$$P_\theta(a) = \frac{1}{N_{\theta,n}} \prod_{i=1}^n e^{-\theta/i} \frac{(\theta/i)^{a_i}}{a_i!}, \quad (2.3.15)$$

with $N_{\theta,n}$ the normalizing constant, we see that P_θ can be interpreted as the distribution of a random vector (A_1, \dots, A_n) whose components A_i are independent and Poisson distributed with mean θ/i *conditioned* on $\sum_{i=1}^n iA_i = n$.

Note that $K_n = \sum_{i=1}^n 1\{A_i > 0\}$, which establishes the link with Section 2.3.1.

2.3.3 GEM-distribution

The next result describes the behavior of Ewens sampling formula in the limit as $n \rightarrow \infty$. Let us call the descendants of the k -th new color in the Hoppe urn model the k -th *family*. This is the collection of all the balls in the urn that carry color k . Let

$$S_k(n) = \text{the size of the } k\text{-th family at time } n. \quad (2.3.16)$$

Consider the random vector

$$S(n) = \left(\frac{1}{n} S_1(n), \frac{1}{n} S_2(n), \frac{1}{n} S_3(n), \dots \right). \quad (2.3.17)$$

Note that the components of this vector sum up to 1.

Theorem 2.3.3 As $n \rightarrow \infty$, $S(n) \Longrightarrow B$, where $B = (B_1, B_2, B_3, \dots)$ is the random vector given by

$$B_k = \left[\prod_{j=1}^{k-1} (1 - Z_j) \right] Z_k, \quad (2.3.18)$$

where (Z_j) are i.i.d. with distribution $\text{Beta}(1, \theta)$, i.e., with density $f(z) = \theta(1 - z)^{\theta-1}$, $z \in [0, 1]$.

Proof. We make use of the following observation, relating the Hoppe urn model to a branching process with immigration:

- Immigrants enter the population at the times of a Poisson process with rate θ . Each individual performs a binary branching process, i.e., gives birth to a new individual at rate 1 (and does not die).

The key observation is:

- If each immigrant is a *new* type and offspring are the *same* type as their parents, then the successive states of the branching process with immigration have the same distribution as the successive states of the Hoppe urn model.

What is done here is that the Hoppe urn model (discrete in time) is *embedded* into the branching process with immigration (continuous in time).

Armed with this observation, we argue as follows. Let $X(t)$ be the number of individuals at time t in the branching process *without* immigration. Then it is well known that

$$e^{-t}X(t) \Longrightarrow \mathcal{E} \quad \text{as } t \rightarrow \infty \quad (2.3.19)$$

with \mathcal{E} an $\text{EXP}(1)$ random variable. Let $X_k(t)$ be the number of individuals in the k -th family at time t in the branching process *with* immigration. Then (2.3.19) implies that

$$e^{-t}(X_1(t), X_2(t), X_3(t), \dots) \Longrightarrow (e^{-T_1}\mathcal{E}_1, e^{-T_2}\mathcal{E}_2, e^{-T_3}\mathcal{E}_3, \dots) \quad \text{as } t \rightarrow \infty, \quad (2.3.20)$$

where \mathcal{E}_k are i.i.d. copies of \mathcal{E} and $0 < T_1 < T_2 < T_3 < \dots$ are the successive arrival times of the Poisson process with rate θ . Let $I(t) = X_1(t) + X_2(t) + X_3(t) + \dots$ be the total number of individuals at time t . Then (2.3.20) gives ($=^{\text{def}}$ means definition)

$$e^{-t}I(t) \Longrightarrow \sum_{k=1}^{\infty} e^{-T_k}\mathcal{E}_k =^{\text{def}} \sigma \quad \text{as } t \rightarrow \infty. \quad (2.3.21)$$

Since $T_k - T_{k-1}$ are i.i.d. $\text{EXP}(\theta)$, it is not hard to show that σ has distribution $\text{GAMMA}(\theta, 1) = \text{EXP}(1/\theta)$.

Next, we note that the limit B_k we are after in (2.3.18) arises as

$$\frac{X_k(t)}{I(t)} \Longrightarrow B_k = \frac{e^{-T_k}\mathcal{E}_k}{\sigma} \quad \text{as } t \rightarrow \infty, \quad (2.3.22)$$

because the embedding of the Hoppe urn model into the branching process with immigration. To identify the distribution of B_k , we consider first the case $k = 1$. We have

$$\frac{e^{-T_1} \mathcal{E}_1}{\sigma} = \frac{\mathcal{E}_1}{\mathcal{E}_1 + \sum_{k=2}^{\infty} e^{-(T_k - T_1)} \mathcal{E}_k} \stackrel{\text{def}}{=} Z_1. \quad (2.3.23)$$

The sum in the denominator is independent of \mathcal{E}_1 and has the same distribution as σ . Hence

$$B_1 = Z_1 \stackrel{\text{dis}}{=} \frac{\mathcal{E}_1}{\mathcal{E}_1 + \sigma} = \text{BETA}(1, \theta), \quad (2.3.24)$$

where $\stackrel{\text{dis}}{=}$ denotes equality in distribution. This proves the claim for $k = 1$. For $k = 2$ we write

$$B_2 = \frac{e^{-T_2} \mathcal{E}_2}{\sigma} = \frac{\sigma}{\mathcal{E}_1 + \sigma} = (1 - Z_1) \frac{e^{-T_2} \mathcal{E}_2}{\sum_{k=2}^{\infty} e^{-T_k} \mathcal{E}_k}. \quad (2.3.25)$$

The last ratio can be rewritten as

$$\frac{\mathcal{E}_2}{\mathcal{E}_2 + \sum_{k=3}^{\infty} e^{-(T_k - T_2)} \mathcal{E}_k} \stackrel{\text{def}}{=} Z_2. \quad (2.3.26)$$

Clearly, Z_2 has the same distribution as Z_1 , and so we get $B_2 = (1 - Z_1)Z_2$. A similar argument applies for $k > 2$. \blacksquare

The limiting distribution in Theorem 2.3.3 is called the Griffiths, Engen, McCloskey distribution (GEM). It describes the frequency spectrum of the types in a very large sample in the *age order*, i.e., the types are ordered according to their first appearance.

2.4 Wright-Fisher with mutation and infinitely many sites

In the infinite alleles model in Section 2.3 we only kept track of when sequences became different through mutation. We did not keep track of the *number* of differences between the sequences. In the present section we will introduce a model that keeps track of where in the sequence the mutations occur.

2.4.1 Infinite sites model

Given are two DNA sequences of length L . Think of these as a sample of size $n = 2$ drawn from a population with $2N$ individuals, each having a type that is given by its sequence of nucleotides of length L . Let Δ_2 be the number of pairwise differences. An example with $L = 14$ is:

Fig. 7 AATCGCTTGATACC
 ACTCGCCTGATAAC

which has pairwise differences at positions 2, 7 and 13, so that $\Delta_2 = 3$. Let $\mu = \mu(N) = \theta/4N$ be the mutation rate as in (2.3.1). We are interested in the distribution of Δ_2 in the limit as $N \rightarrow \infty$ and $L \rightarrow \infty$. The latter limit is why the model is called the *infinite sites model*.

Lemma 2.4.1 Δ_2 has distribution $\text{GEO}(1/(\theta + 1))$, i.e.,

$$\mathbb{P}(\Delta_2 = k) = \frac{1}{\theta + 1} \left(\frac{\theta}{\theta + 1} \right)^k, \quad k \in \mathbb{N}_0. \quad (2.4.1)$$

Proof. In the limit as $L \rightarrow \infty$, mutations always occur at different sites. The probability of *coalescence before mutation* is $1/(\theta + 1)$ (in the limit as $N \rightarrow \infty$). If mutation occurs before coalescence, which happens with probability $\theta/(\theta + 1)$, then there is an equal chance of having another mutation before coalescence, i.e., the system starts from scratch. Hence, the probability to have k mutations followed by coalescence is given by the right-hand side of (2.4.1). ■

It follows from Lemma 2.4.1 that

$$\mathbb{E}(\Delta_2) = \theta, \quad \text{Var}(\Delta_2) = \theta(1 + \theta). \quad (2.4.2)$$

As mentioned at the end of Section 2.3.1, it is of interest to estimate θ from data. Therefore we ask the same question for a sample of size n ($\ll N \rightarrow \infty$) drawn from the population. To that end, given n DNA sequences, let Δ_{ij} be the number of pairwise differences between the i -th and the j -th sequence. Put

$$\Delta_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \Delta_{ij}. \quad (2.4.3)$$

Lemma 2.4.2 For $n \geq 2$,

$$\mathbb{E}(\Delta_n) = \theta, \quad \text{Var}(\Delta) = \frac{n+1}{3(n-1)}\theta + \frac{2(n^2+n+3)}{9n(n-1)}\theta^2. \quad (2.4.4)$$

Proof. The first claim is immediate from the fact that $\mathbb{E}(\Delta_{ij}) = \theta$ for all i, j , as is obvious from exchangeability and the first half of (2.4.2). The second claim cannot be deduced from the second half of (2.4.2) because the Δ_{ij} are not independent. In fact, we need to compute three different variances, as we show next.

Write

$$[\Delta_n]^2 = \binom{n}{2}^{-2} \sum_{1 \leq i < j \leq n} \Delta_{ij} \sum_{1 \leq k < l \leq n} \Delta_{kl}. \quad (2.4.5)$$

There are three types of terms:

- (1) $\binom{n}{2}$ terms with $i = k, j = l$.
- (2) $\binom{n}{2} 2(n-2)$ terms with $i = k, j \neq l$ or $i \neq k, j = l$.
- (3) $\binom{n}{2} \binom{n-2}{2}$ terms with $i \neq k, j \neq l$.

Hence, using exchangeability, we have

$$\begin{aligned}\text{Var}(\Delta_n) &= \binom{n}{2}^{-2} \sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} \text{Cov}(\Delta_{ij}, \Delta_{kl}) \\ &= \binom{n}{2}^{-1} \left[U_2 + 2(n-2)U_3 + \binom{n-2}{2} U_4 \right]\end{aligned}\tag{2.4.6}$$

with

$$\begin{aligned}U_2 &= \text{Var}(\Delta_{12}), \\ U_3 &= \text{Cov}(\Delta_{12}, \Delta_{13}), \\ U_4 &= \text{Cov}(\Delta_{12}, \Delta_{34}).\end{aligned}\tag{2.4.7}$$

Thus, it remains to compute U_2, U_3, U_4 . In other words, general n reduces to $n = 4$.

We already know that $U_2 = \theta + \theta^2$. A somewhat lengthy calculation (see Durrett [4], pp. 45–49) yields

$$\begin{aligned}U_3 &= \frac{1}{2}\theta + \frac{1}{3}\theta^2, \\ U_4 &= \frac{1}{3}\theta + \frac{2}{9}\theta^2.\end{aligned}\tag{2.4.8}$$

After substituting this into (2.4.6), we get

$$\text{Var}(\Delta_n) = \binom{n}{2}^{-1} \left[\frac{1}{6}n(n+1)\theta + \frac{1}{9}(n^2 + n + 3)\theta^2 \right],\tag{2.4.9}$$

which proves the second claim in (2.4.4). ■

Lemma 2.4.2 says that Δ_n is an unbiased estimator of θ . However, it is not asymptotically sharp, because $\text{Var}(\Delta_n) \rightarrow U_4 \neq 0$ as $n \rightarrow \infty$. The fact that $U_4 \neq 0$ is due to coalescence: even though mutations occur independently in the sequences labelled 1, 2, 3, 4, coalescence between these sequences creates a dependence between Δ_{12} and Δ_{34} (which are the pairwise differences between 1, 2, respectively, 3, 4). This is precisely why the computation of U_3, U_4 is somewhat delicate.

Despite the lack of asymptotic sharpness of Δ_n as $n \rightarrow \infty$, Lemma 2.4.2 still allows for an estimate of θ with some confidence interval. See Durrett [4], pp. 43–44.

2.4.2 Segregating sites

We next look at the number of *segregating* sites, i.e.,

$$S_n = \text{the number of sites where some pair in the } n \text{ sample differs.}\tag{2.4.10}$$

Lemma 2.4.3 $\mathbb{E}(S_n) = \theta \sum_{i=1}^{n-1} \frac{1}{i} \sim \theta \log n$ as $n \rightarrow \infty$.

Proof. For $j \geq 2$, let t_j be the amount of time in the coalescent during which there are j lineages. In Section 2.1 we saw that, if N is large and time is measured in units of $2N$

generations, then t_j is exponential with mean $1/[\frac{1}{2}j(j-1)]$. The total time associated with the coalescent (all times along all branches), starting from the n individuals in the sample, is

$$T = \sum_{j=2}^n j t_j. \quad (2.4.11)$$

From this we get

$$\mathbb{E}(T) = \sum_{j=2}^n j \mathbb{E}(t_j) = \sum_{j=2}^n j \binom{j}{2}^{-1} = \sum_{j=2}^n 2 \frac{2}{j-1} = 2 \sum_{i=1}^{n-1} \frac{1}{i}. \quad (2.4.12)$$

Since μ is the mutation rate per locus, after time is scaled up by $2N$ the mutation rate per locus is $2N\mu = \frac{1}{2}\theta$. Hence

$$\mathbb{E}(S_n) = \frac{1}{2}\theta \mathbb{E}(T), \quad (2.4.13)$$

since a mutation can occur anywhere on the coalescent and each mutation contributes 1 to the number of segregating sites. ■

Lemma 2.4.3 shows that $S_n / \sum_{i=1}^n \frac{1}{i}$ is an unbiased estimator of θ . What about its variance?

Lemma 2.4.4 $\text{Var}(S_n) = \theta \sum_{i=1}^n \frac{1}{i} + \theta^2 \sum_{i=1}^n \frac{1}{i^2} \sim \theta \log n$ as $n \rightarrow \infty$.

Proof. Let s_j be the number of segregating sites created when there are j lineages. Since mutations occur at rate $\frac{1}{2}\theta$ and coalescence occurs at rate $\frac{1}{2}j(j-1)$, we have

$$\mathbb{P}(s_j = k) = \frac{j-1}{\theta + j - 1} \left(\frac{\theta}{\theta + j - 1} \right)^k, \quad k \in \mathbb{N}. \quad (2.4.14)$$

Hence,

$$\text{Var}(s_j) = \frac{1}{j-1} \theta + \frac{1}{(j-1)^2} \theta^2. \quad (2.4.15)$$

Since $\text{Var}(S_n) = \sum_{j=2}^n \text{Var}(s_j)$, we get the claim after putting $i = j - 1$. ■

Lemma 2.4.4 shows that the statistical estimation of θ via the quantity S_n suffers from the same slowness as via the quantity K_n in Lemma 2.3.1: the variance of the estimator decays like $1/\sqrt{\log n}$.

The exact distribution of S_n can be computed iteratively from the recursion relation $S_n = S_{n-1} + s_n$ and the distribution of s_n given in (2.4.14), using the fact that s_n is independent of S_{n-1} . For instance,

$$\mathbb{P}(S_n = 0) = \prod_{j=2}^n \mathbb{P}(s_j = 0) = \prod_{j=2}^n \frac{j-1}{\theta + j - 1} = \frac{(n-1)!}{(\theta+1) \times \cdots \times (\theta+n-1)}. \quad (2.4.16)$$

This formula is reminiscent of Ewens sampling formula in (2.3.10).

2.5 Wright-Fisher with selection

In all the models considered so far the evolution of the population was *neutral*: all types were equally fit for reproduction. In this section we look at the two-type WF-model with *selection*: one type reproduces faster than the other. We will pass to continuous time right away, so we start from the Moran-model introduced in Section 2.1.4.

2.5.1 Selection

Our model is the birth-death process on state space $\Omega = \{0, 1, \dots, 2N\}$ with transition rates

$$\begin{aligned} i \rightarrow i + 1 \text{ at rate } b_i &= (2N - i) \frac{i}{2N}, \\ i \rightarrow i - 1 \text{ at rate } d_i &= i \frac{2N - i}{2N} (1 - s), \end{aligned} \tag{2.5.1}$$

where $s \in (0, 1)$ is a *selection parameter*. Here, as before, individuals randomly choose an ancestor and adopts its type, but type a does so at rate 1 and type A at rate $1 - s$. Thus, type A evolves more slowly than type a . The *relative fitness* of the two types is $1 - s$.

Let X_t denote the state of the system at time t , starting from a given X_0 . As before, the states $i = 0$ (all a) and $i = 2N$ (all A) are traps, so that eventually type fixation will occur. We are interested in how the parameter s influences the probability of fixation in these two traps as well as the fixation time.

We saw in Lemma 2.1.1 that in the neutral case $s = 0$ the probability of fixation at $2N$ given $X_0 = i$ equals $\frac{i}{2N}$.

Lemma 2.5.1 *Let $s \in (0, 1)$. Let $\tau = \min\{t \geq 0: X_t = 0 \text{ or } X_t = 2N\}$. Then*

$$\mathbb{P}(X_\tau = 2N \mid X_0 = i) = \frac{1 - (1 - s)^i}{1 - (1 - s)^{2N}}. \tag{2.5.2}$$

Proof. Abbreviate $g(i) = \mathbb{P}(X_\tau = 2N \mid X_0 = i)$. By considering what happens at the first transition away from i , we have

$$g(i) = \frac{b_i}{b_i + d_i} g(i + 1) + \frac{d_i}{b_i + d_i} g(i - 1), \quad \Omega \setminus \{0, 2N\}. \tag{2.5.3}$$

This recursion is to be solved subject to the boundary conditions $g(0) = 0$ and $g(2N) = 1$. Rearranging the terms in (2.5.3), we may write

$$[g(i + 1) - g(i)] = \frac{d_i}{b_i} [g(i) - g(i - 1)] = (1 - s) [g(i) - g(i - 1)]. \tag{2.5.4}$$

Use $g(0) = 0$, put $c = g(1)$ and iterate (2.5.4), to get

$$g(i + 1) - g(i) = c(1 - s)^i. \tag{2.5.5}$$

Summing this over $i = 0, \dots, j - 1$, we obtain

$$g(j) = \sum_{i=0}^{j-1} c(1-s)^i = \frac{c}{s} [1 - (1-s)^j], \quad j \in \Omega. \quad (2.5.6)$$

Since $g(2N) = 1$, we have

$$c = s [1 - (1-s)^{2N}]^{-1}, \quad (2.5.7)$$

which proves the claim. \blacksquare

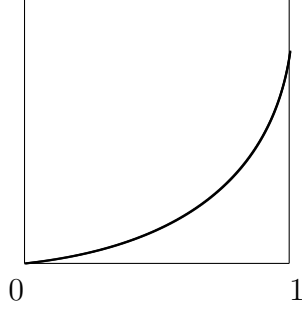


Fig. 8. Qualitative picture of (2.5.2).

The average fixation time $\mathbb{E}(\tau)$ is not so easy to compute. Even for the WF-model ($s = 0$) we did not find a closed formula, but we did see that it is of order N (recall Lemma 2.1.2 and (2.1.11)).

Lemma 2.5.2 *Let $s \in (0, 1)$. Then*

$$\mathbb{E}(\tau \mid X_0 = 1, X_\tau = 2N) \sim \frac{2}{s} \log N \quad \text{as } N \rightarrow \infty. \quad (2.5.8)$$

Proof. The proof is given in Durrett [4], pp. 119-123. It is based on a somewhat lengthy calculation showing that it takes time

- (1) $\sim \frac{1}{s} \log N$ to go from $i = 1$ to $i = \frac{2N}{\log N}$;
- (2) $\sim 2 \log \log N$ to go from $i = \frac{2N}{\log N}$ to $i = 2N[1 - \frac{1}{\log N}]$;
- (3) $\sim \frac{1}{s} \log N$ to go from $i = 2N[1 - \frac{1}{\log N}]$ to $i = 2N$.

Thus, most of the time is spent near the boundaries of Ω , where one of the types dominates, and the transition from the left to the right boundary is very fast, essentially deterministic. \blacksquare

The fast crossing from $i = \frac{2N}{\log N}$ to $i = 2N[1 - \frac{1}{\log N}]$ is called a *selective sweep*. Note that the slower type A has a selective advantage over the faster type a . The reason is that the faster type more rapidly adopts the slower type.

2.5.2 Weak selection

Interesting behavior occurs when there is *weak selection*, in particular, when $N \rightarrow \infty$ and

$$s = s(N) = \frac{\sigma}{4N}, \quad \sigma \in (0, \infty). \quad (2.5.9)$$

Think of σ as the *fitness advantage of the population*. Write $X_t^{(N)}$ to exhibit the N -dependence and define, in analogy with (2.1.12),

$$Y_t^{(N)} = \frac{1}{2N} X_{\lceil 2Nt \rceil}^{(N)}, \quad t \geq 0. \quad (2.5.10)$$

The following result is the analogue of Theorem 2.1.3. We assume that

$$Y_0^{(N)} \implies Y_0 \quad \text{as } N \rightarrow \infty \quad (2.5.11)$$

and ask whether

$$(Y_t^{(N)})_{t \geq 0} \implies (Y_t)_{t \geq 0} \quad \text{as } N \rightarrow \infty. \quad (2.5.12)$$

Theorem 2.5.3 *The scaling in (2.5.11) subject to (2.5.12) holds true, with $(Y_t)_{t \geq 0}$ the diffusion process on $[0, 1]$ given by the stochastic differential equation*

$$dY_t = \sigma Y_t(1 - Y_t) dt + \sqrt{Y_t(1 - Y_t)} dW_t. \quad (2.5.13)$$

Proof. The proof is informal. Both $(Y_t^{(N)})_{t \geq 0}$ and $(W_t)_{t \geq 0}$ are Markov processes. Compute, for $y \in [0, 1]$ and small $\Delta t > 0$,

$$\begin{aligned} \mathbb{E} \left(Y_{t+\Delta t}^{(N)} - Y_t^{(N)} \mid Y_t^{(N)} = y \right) &= \sigma y(1 - y)\Delta t + o(\Delta t), \\ \mathbb{E} \left(\left[Y_{t+\Delta t}^{(N)} - Y_t^{(N)} \right]^2 \mid Y_t^{(N)} = y \right) &= y(1 - y)\Delta t + o(\Delta t). \end{aligned} \quad (2.5.14)$$

The first line comes from the observation that

$$b_{\lceil 2Ny \rceil} - d_{\lceil 2Ny \rceil} \rightarrow 2Ny(1 - y)[1 - (1 - s)] = 2Ns y(1 - y) = \sigma y(1 - y) \quad \text{as } N \rightarrow \infty, \quad (2.5.15)$$

where the scalings in space and time cancel each other. The second line is the same calculation as for $s = 0$. ■

The stochastic differential equation in (2.5.13) has two parts: a *deterministic* part, given by a logistic drift term pushing the system towards all A , and a *random* part, given by the WF-diffusion term. Like (2.1.15), it has a unique strong solution. A dual process exists, but this is more complex than the death process we encountered in Theorem 2.1.4 as the dual process for the WF-diffusion ($\sigma = 0$). This dual process requires a graphical construction (see Durrett [4], pp. 127–128). We see from (2.5.10–2.5.12) and Theorem 2.5.3 that under weak selection the fixation time is of order N .

Let us next see what happens when we add *mutation*. Then the states $i = 0$ and $i = 2N$ are no longer traps and the system has a non-trivial equilibrium. We consider the limit $N \rightarrow \infty$ with

$$u = u(N) = \frac{q}{4N}, \quad v = v(N) = \frac{r}{4N}, \quad s = s(N) = \frac{\sigma}{4N}. \quad (2.5.16)$$

The following generalizes Theorem 2.2.2.

Theorem 2.5.4 *Subject to (2.5.16), the convergence in (2.2.15) holds true, with Y the random variable on $[0, 1]$ with probability density*

$$f(x) = C_{q,r,\sigma} x^{q-1} (1-x)^{r-1} \exp\left[\frac{1}{2}\sigma x\right], \quad x \in [0, 1] \quad (2.5.17)$$

with $C_{q,r,\sigma}$ the normalizing constant.

Proof. The proof for $\sigma = 0$ was based on an asymptotic expansion (recall (2.2.20–2.2.23)). The proof for $\sigma > 0$ can be done quickly by comparison with the case $\sigma = 0$. To that end, let $\pi(i)$, $i \in \Omega$, be the stationary distribution before passing to the limit $N \rightarrow \infty$. Let b_i^* , d_i^* be the birth and death rates for the model *without* selection, given by (2.1.26). Then

$$\pi(i)d_i = \pi(i-1)b_{i-1}, \quad i \in \Omega \setminus \{0\}. \quad (2.5.18)$$

Hence, for any $k < i$, we have

$$\pi(i) = \pi(k) \left[\prod_{j=k+1}^i \frac{b_{j-1}}{d_j} \right] = \pi(k) b_k \left[\prod_{j=k+1}^{i-1} \frac{b_j}{d_j} \right] \frac{1}{d_i}. \quad (2.5.19)$$

Rewrite the right-hand side as

$$\frac{1}{(1-s)^{i-k}} \pi(k) b_k^* \left[\prod_{j=k+1}^{i-1} \frac{b_j^*}{d_j^*} \right] \frac{1}{d_i^*}, \quad (2.5.20)$$

where the factor $1/(1-s)^{i-k}$ arises from the fact that $b_j = b_j^*$ and $d_j = (1-s)d_j^*$. Pick $k = N$ and $i = \lceil 2Nx \rceil$, and let $N \rightarrow \infty$. Then

$$(1-s)^{-(i-k)} \sim \left(1 - \frac{\sigma}{4N}\right)^{-2N(x-\frac{1}{2})} \sim \exp\left[\frac{1}{2}\sigma\left(x - \frac{1}{2}\right)\right]. \quad (2.5.21)$$

This explains why the extra factor $\exp[\frac{1}{2}\sigma x]$ shows up in (2.5.17) in comparison with (2.2.16). The choice $k = N$ is arbitrary. This only serves as a reference point for the computation and cancels out after normalization. \blacksquare

3 Interacting Wright-Fisher populations

In Chapter 2 we considered models where a *single* population evolves under *random resampling, mutation and selection*. In the present chapter we move on towards models with *several interacting populations*.

We imagine that each population lives in a *colony* and that different colonies are organized into a *lattice*, \mathcal{L} , which labels the colonies. In what follows we will consider two choices:

- (1) $\mathcal{L} = \mathbb{Z}^2$ (Section 3.1),
- (2) $\mathcal{L} = \Omega_M$ (Section 3.2), the hierarchical group of order $M \in \mathbb{N}$.

We allow the individuals to choose their ancestors not only from their *own* colony, but also from *other* colonies, according to a prescribed transition kernel $p(x, y)$, $x, y \in \mathcal{L}$. This is called *migration*, because it causes the types to migrate from one colony to another. Note that the individuals themselves do not migrate. In each colony the population remains fixed at $2N$ individuals.

3.1 The stepping stone model

In this section we consider interacting WF-models with mutation labelled by the square lattice

$$\mathcal{L} = \mathbb{Z}^2 \tag{3.1.1}$$

and subject to migration with transition kernel

$$p(x, y) = (1 - \nu)\delta_{x,y} + \nu q(y - x), \quad x, y \in \mathbb{Z}^2. \tag{3.1.2}$$

where $q: \mathbb{Z}^2 \rightarrow [0, 1]$ is a prescribed probability distribution on \mathbb{Z}^2 and $\nu \in (0, 1)$ is a parameter.

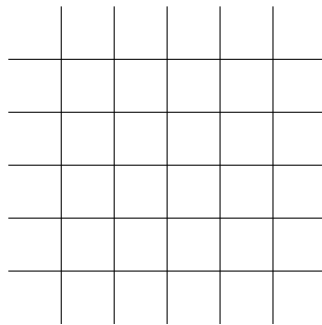


Fig. 9. The square lattice \mathbb{Z}^2 .

The evolution of the system is as follows:

- (1) At each time step, each individual mutates to a new type with probability μ .

- (2) At each time step, each individual randomly chooses a colony (migration) and adopts the type of a randomly chosen individual from that colony. Individuals in colony x choose colony y with probability $(1 - \mu)p(x, y)$.
- (3) Each colony contains $2N$ individuals. All individuals mutate and migrate independently.

The migration under (3.1.2) amount to individuals choosing their own colony with probability ν and another colony with probability $1 - \nu$. Given that they choose another colony, the choice is made with a probability that only depends on the distance to this colony. The parameter ν is the *migration probability*.

In what follows we will focus on the nearest-neighbor model:

$$q(z) = \begin{cases} \frac{1}{4} & \text{if } \|x\| = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (3.1.3)$$

Most of the results to be described below carry over to more general q , provided it is symmetric and has finite support.

3.1.1 Lineages

We are interested in the lineage of two individuals randomly drawn from colonies x and y when the system is in equilibrium. When $x = y$, we assume that two *distinct* individuals are drawn from the colony. Due to the presence of mutations, there are no traps and the system has an ergodic equilibrium.

Let

$$\psi(x, y) = \text{probability that two individuals randomly drawn from colonies } x \text{ and } y \text{ are identical by descent in equilibrium.} \quad (3.1.4)$$

We will compute this quantity and study its dependence on x and y . By translation invariance, $\psi(x, y)$ is a function of $y - x$ only.

We begin by expressing $\psi(x, y)$ in terms of $\psi(0, 0)$ and the iterates of the transition kernel $p(x, y)$.

Lemma 3.1.1 *For all $x, y \in \mathbb{Z}^2$,*

$$\psi(x, y) = \frac{1 - \psi(0, 0)}{2N} \sum_{n=1}^{\infty} (1 - \mu)^{2n} p^{2n}(x, y). \quad (3.1.5)$$

Proof. Let $\psi_n(x, y)$ be the required probability at time n when the system starts from some given initial configuration. Then we have the recursion relation

$$\begin{aligned} \psi_{n+1}(x, y) = & (1 - \mu)^2 \sum_{z \in \mathbb{Z}^2} p(x, z)p(y, z) \frac{1 - \psi_n(z, z)}{2N} \\ & + (1 - \mu)^2 \sum_{x', y' \in \mathbb{Z}^2} p(x, x')p(y, y') \psi_n(x', y'). \end{aligned} \quad (3.1.6)$$

Here, the second term sums over the two backward steps $x \rightarrow x'$ and $y \rightarrow y'$ in the lineages of the two individuals and requires that no mutation occurs during either of these steps, after which the system continues from x' and y' . The first term compensates for the two lineages choosing the same parent when they choose the same colony (because in our definition of $\psi(x, y)$ and $\psi_n(x, y)$ for $x = y$ we insisted on drawing two distinct individuals from the colony). In equilibrium we have $\psi_{n+1} = \psi_n = \psi$, and so we get

$$\psi(x, y) = (1 - \mu)^2 \sum_{z \in \mathbb{Z}^2} p(x, z)p(y, z) \frac{1 - \psi(z, z)}{2N} + (1 - \mu)^2 \sum_{x', y' \in \mathbb{Z}^2} p(x, x')p(y, y') \psi(x', y'). \quad (3.1.7)$$

The first term in (3.1.7) gives the first summand in (3.1.5), because $\psi(z, z) = \psi(0, 0)$ and $\sum_z p(x, z)p(y, z) = \sum_z p(x, z)p(z, y) = p^2(x, y)$. The second term can be iterated by resubstitution of (3.1.7). This generates all the higher summands in (3.1.5). Convergence of the infinite series is no problem because $\mu > 0$. ■

Abbreviate

$$G_{\mu, \nu}(x, y) = \sum_{n=1}^{\infty} (1 - \mu)^{2n} p^{2n}(x, y). \quad (3.1.8)$$

Here, we add the lower index ν because the transition kernel p depends on ν .

Theorem 3.1.2 For $\mu, \nu \in (0, 1)$ and $x, y \in \mathbb{Z}^2$,

$$\psi(x, y) = \frac{G_{\mu, \nu}(x, y)}{2N + G_{\mu, \nu}(0, 0)}. \quad (3.1.9)$$

Proof. By picking $x = y = 0$ in (3.1.5) and solving for $\psi(0, 0)$, we get

$$\psi(0, 0) = \frac{G_{\mu, \nu}(0, 0)}{2N + G_{\mu, \nu}(0, 0)}. \quad (3.1.10)$$

Substituting this into (3.1.5), we get the claim. ■

Thus, all that we need to know is how $G_{\mu, \nu}(x, y)$ behaves as a function of $y - x$ and μ, ν .

Durrett [4], pp. 104–108, provides a Fourier calculation of $G_{\mu, \nu}(x, y)$ for the case where the lattice $\mathcal{L} = \mathbb{Z}^2$ is restricted to a finite box of size $L \in \mathbb{N}$,

$$\mathcal{L} = [0, L]^2 \cap \mathbb{Z}^2, \quad (3.1.11)$$

with *periodic boundary conditions*, turning the box into a torus. For this choice there are two regimes:

$$\begin{aligned} \text{I.} & \quad \frac{1}{\mu} \ll \frac{1}{\nu} L^2, \\ \text{II.} & \quad \frac{1}{\mu} \gg \frac{1}{\nu} L^2. \end{aligned} \quad (3.1.12)$$

Regime I corresponds to the system *not feeling the boundary of the box*. Indeed, $\frac{1}{\mu}$ is the average time required for a mutation to occur in one lineage. Consequently, $\frac{1}{\nu}$ is

the variance of the displacement of a lineage before a mutation occurs (the variance per step is 1). If this ratio is much smaller than L^2 , then a mutation will occur before the lineages notice the boundary of the box, and the behavior of $\psi(x, y)$ will be essentially independent of L . Regime II, on the other hand, corresponds to the situation where the lineages *wrap themselves around the torus many times* before a mutation occurs. In that case, the behavior of $\psi(x, y)$ will depend on L .

3.1.2 Regime I

Theorem 3.1.3 Write $\psi(x, y) = \phi(y - x)$. In Regime I,

$$\phi(z) \asymp \frac{2}{8\pi\nu N + \log(1/2\mu)} \left[K_0 \left(\sqrt{\frac{2\mu}{\nu}} \|z\| \right) - K_0 \left(\sqrt{\frac{1}{\nu}} \|z\| \right) \right] \quad (3.1.13)$$

with K_0 a Bessel function satisfying

$$K_0(u) \sim \begin{cases} \log(1/u) & \text{for } u \downarrow 0, \\ \sqrt{\pi/2u} e^{-u} & \text{for } u \uparrow \infty. \end{cases} \quad (3.1.14)$$

Proof. See Durrett [4], p. 107. ■

The result in Theorem 3.1.3 shows how, on a sufficiently large box, the probability for two random individuals to be identical by descent depends on their distance. The answer depends on all three parameters N, μ, ν , subject to the constraint $L^2 \gg \nu/\mu$.

3.1.3 Regime II

Regime II is harder to come by. Our first intuition is that the system behaves like a “homogeneously mixing population”, because the lineages wrap themselves around the torus many times before a mutation occurs and therefore sample the full torus. In other words, the system behaves like a WF-model with a single colony containing $2NL^2$ individuals, so-called *panmictic* behavior. However, it turns out that this intuition is not quite correct. In fact, we will see that this behavior only shows up when

$$\mu \ll 1 \quad \text{and} \quad N\nu \gg \log L, \quad (3.1.15)$$

- Regime II without mutation:

To investigate Regime II, we first consider the case $\mu = 0$. Since mutations occur independently of everything else, it will be easy to incorporate them later. Let

$$T_0 = \text{the time required for two lineages to meet in the same colony.} \quad (3.1.16)$$

Only after meeting in the same colony can the two lineages coalesce. However, to do so they must also choose the same ancestor in this colony. We will investigate that aspect later. We write \mathbb{P}_π for the law of the two lineages when they start from two *randomly* chosen locations in the box.

Lemma 3.1.4 For $\mu = 0$ and $\nu \in (0, 1]$,

$$\lim_{L \rightarrow \infty} \mathbb{P}_\pi \left(T_0 > \frac{L^2 \log L}{2\pi\nu} t \right) = e^{-t} \quad \forall t \geq 0. \quad (3.1.17)$$

Proof. The order of T_0 can be seen from the following argument. When read backwards, the two lineages perform nearest-neighbor random walks. Let X_n be the distance between these two random walks after time n (read backwards). Since the two lineages start from two randomly chosen sites of the torus, we have $\mathbb{P}_\pi(X_n = 0) = 1/L^2$ for all n . Therefore we can write

$$\begin{aligned} 1 &= \sum_{n=0}^{L^2-1} \mathbb{P}_\pi(X_n = 0) \\ &= \sum_{m=0}^{L^2-1} \mathbb{P}_\pi(T_0 = m) \sum_{l=m}^{L^2-1} \mathbb{P}(X_l = 0 \mid X_m = 0) \\ &= \sum_{m=0}^{L^2-1} \mathbb{P}_\pi(T_0 = m) \sum_{k=0}^{L^2-1-m} \mathbb{P}(X_k = 0 \mid X_0 = 0), \end{aligned} \quad (3.1.18)$$

where the second sum counts the meetings of the two random walks following their first meeting. By the local central limit theorem for nearest-neighbor random walk on \mathbb{Z}^2 , we have

$$\mathbb{P}(X_k = 0 \mid X_0 = 0) \sim \frac{1}{2\pi} \frac{1}{2\nu k} \quad \text{as } k \rightarrow \infty, \quad (3.1.19)$$

because the random walks each make $\sim \nu k$ steps until time k . Inserting (3.1.19) into (3.1.18), we obtain

$$1 \sim \mathbb{P}_\pi(T_0 \leq L^2) \frac{\log L^2}{2\pi(2\nu)}, \quad (3.1.20)$$

or

$$\mathbb{P}_\pi(T_0 \leq L^2) \sim \frac{2\pi\nu}{\log L}. \quad (3.1.21)$$

Thus, on the order of $\log L/(2\pi\nu)$ time intervals of length L^2 are needed before the two random walks meet, which means that T_0 is of order $L^2 \log L/(2\pi\nu)$, as claimed.

To get the exponential distribution, we need to show that the difference random walk reaches equilibrium in a time $o(L^2 \log L/(2\pi\nu))$, since this is what provides the “lack of memory” property of the exponential. The proof of this fact is more technical and we skip the argument. See Cox and Durrett [2]. \blacksquare

A more refined result is obtained by looking at two lineages that start not randomly but at fixed colonies, say 0 and x . We write $\mathbb{P}_{x,0}$ for their law.

Lemma 3.1.5 For $\mu = 0$, $\nu \in (0, 1]$ and $\beta \in [0, 1]$,

$$\lim_{L \rightarrow \infty} \mathbb{P}_{x(L),0} \left(T_0 > \frac{L^2 \log L}{2\pi\nu} t \right) = \beta e^{-t} \quad \forall t \geq 0, \quad (3.1.22)$$

for any $x(L)$ such that

$$\lim_{L \rightarrow \infty} \frac{\log \|x(L)\|}{\log L} = \beta. \quad (3.1.23)$$

Proof. The main idea behind this result is the analogue of (3.1.21),

$$\lim_{L \rightarrow \infty} \mathbb{P}_{x(L),0} \left(T_0 \leq \frac{L^2}{\nu} \right) = 1 - \beta, \quad (3.1.24)$$

together with the fact that when $T_0 > L^2/\nu$ it is very likely that $T_0 > L^2\sqrt{\log L}/\nu$ (i.e., either a return occurs quickly or else it takes somewhat longer). The rest of the argument is similar, and we again refer to the literature (see Durrett [4], p. 110). ■

Lemma 3.1.5 says that if the lineages start at distance $|x(L)| \asymp L^\beta$, then there is sharp control over the time when they first meet. Note that all scales $\beta \in [0, 1]$ are relevant. Lemma 3.1.4 corresponds to $\beta = 1$.

Getting the two lineages to the same colony at time T_0 is the first step towards coalescence. Afterwards, either coalescence occurs at that colony (when the lineages choose the same ancestor in that colony) or the lineages separate and they must meet at some later time. The time until coalescence occurs equals

$$t_0 = T_0 + T_1 + T_2 + \dots + T_\sigma, \quad (3.1.25)$$

where T_1, T_2, \dots are i.i.d. copies of the return times of the two lineages (independent of T_0), and σ is an independent geometrically distributed random variable with mean $2N$. The reason is that upon each return the two lineages have probability $\frac{1}{2N}$ to coalesce.

Lemma 3.1.6 For $\mu = 0$ and $\nu \in (0, 1]$,

$$\mathbb{E}_0(T_1) = L^2, \quad \mathbb{E}(\sigma) = 2N. \quad (3.1.26)$$

Proof. The second claim is obvious. To get the first claim, use that the uniform distribution on the torus is the equilibrium distribution. This claim is a well-known mean recurrence time property of Markov chains, for which we refer to standard textbooks. ■

Note that the result in Lemma 3.1.6 is independent of ν . Increasing the pausing probability of the two random walks does not effect their average return time: although an immediate return through pausing is more likely, when the random walks separate it takes longer for them to achieve a return.

A comparison of Lemmas 3.1.4 and 3.1.6 shows that there are two subregimes:

$$\begin{aligned} \text{A. } & \frac{1}{\nu} \log L \ll N : \mathbb{E}(T_0) \ll \mathbb{E}_0(t_0), \\ \text{B. } & \frac{1}{\nu} \log L \gg N : \mathbb{E}(T_0) \gg \mathbb{E}_0(t_0). \end{aligned} \quad (3.1.27)$$

In regime A, the starting locations of the two individuals do not matter, while in regime B they do. For regime A we have:

Lemma 3.1.7 For $\mu = 0$ and $\nu \in (0, 1]$,

$$\lim_{L \rightarrow \infty} \sup_{x \in [0, L]^2 \cap \mathbb{Z}^2} |\mathbb{P}_{x,0}(t_0 > 2NL^2t) - e^{-t}| \rightarrow 0 \quad \forall t \geq 0, \quad (3.1.28)$$

provided $N\nu/\log L \rightarrow \infty$.

Proof. This is essentially immediate from (3.1.25) and Lemmas 3.1.5–3.1.6. \blacksquare

For regime B we have:

Lemma 3.1.8 For $\mu = 0$ and $\nu \in (0, 1]$,

(i) If $\lim_{L \rightarrow \infty} 2\pi N\nu/\log L = \alpha \in [0, \infty)$, then

$$\lim_{L \rightarrow \infty} \mathbb{P}_\pi \left(t_0 > (1 + \alpha) \frac{L^2 \log L}{2\pi\nu} t \right) = e^{-t} \quad \forall t \geq 0. \quad (3.1.29)$$

(ii) If $\lim_{L \rightarrow \infty} 2\pi N\nu/\log L = \alpha \in [0, \infty)$ and $\lim_{L \rightarrow \infty} \log \|x(L)\|/\log L = \beta \in [0, 1]$, then

$$\lim_{L \rightarrow \infty} \mathbb{P}_{X(L),0} \left(t_0 > (1 + \alpha) \frac{L^2 \log L}{2\pi\nu} t \right) = \left(\beta + (1 - \beta) \frac{\alpha}{1 + \alpha} \right) e^{-t} \quad \forall t \geq 0. \quad (3.1.30)$$

Proof. This is again essentially immediate from (3.1.25) and Lemmas 3.1.5–3.1.6. \blacksquare

Let us summarize the above observations.

Theorem 3.1.9 If

$$\lim_{L \rightarrow \infty} \frac{2\pi N\nu}{\log L} = \alpha \in [0, \infty], \quad \lim_{L \rightarrow \infty} \frac{\log \|x(L)\|}{\log L} = \beta \in [0, 1], \quad (3.1.31)$$

then under \mathbb{P}_π ,

$$\begin{aligned} \alpha = \infty : & \quad \frac{t_0}{c_L} \implies \text{EXP}(1) \quad \text{with } C_L = 2NL^2, \\ \alpha \in [0, \infty) : & \quad \frac{t_0}{c_L} \implies \text{EXP}(1) \quad \text{with } C_L = (1 + \alpha) \frac{L^2 \log L}{2\pi\nu}, \end{aligned} \quad (3.1.32)$$

while under $\mathbb{P}_{x(L),0}$,

$$\begin{aligned} \alpha = \infty : & \quad \frac{t_0}{c_L} \implies \text{EXP}(1) \quad \text{with } C_L = 2NL^2, \\ \alpha \in [0, \infty) : & \quad \frac{t_0}{c_L} \implies (1 - \gamma)\delta_0 + \gamma\text{EXP}(1) \quad \text{with } C_L = (1 + \alpha) \frac{L^2 \log L}{2\pi\nu}, \end{aligned} \quad (3.1.33)$$

where

$$\gamma = \beta + (1 - \beta) \frac{\alpha}{1 + \alpha}. \quad (3.1.34)$$

Proof. See Lemmas 3.1.7 and 3.1.8. ■

• Regime II with mutation:

Having thus obtained the asymptotics of the time to coalescence, we reintroduce mutation and return to the probability of the two lineages being equal by descent, defined in (3.1.4). Define

$$\begin{aligned} h(x) &= \mathbb{E}_{x,0}([1 - \mu]^{2t_0}), \\ h &= \mathbb{E}_\pi([1 - \mu]^{2t_0}). \end{aligned} \quad (3.1.35)$$

Then

$$\begin{aligned} h(x) &= \psi(x, 0), \\ h &= \sum_{x,y \in [0,L]^2 \cap \mathbb{Z}^2} \psi(x, y). \end{aligned} \quad (3.1.36)$$

Inserting $t_0/C_L \implies (1 - \gamma) + \gamma \text{EXP}(1)$ into (3.1.35), we compute (with \mathbb{E} standing for either \mathbb{E}_π or $\mathbb{E}_{x,0}$)

$$\begin{aligned} \mathbb{E}([1 - \mu]^{2t_0}) &= \int_0^\infty [1 - \mu]^{2t} \mathbb{P}(t_0 \in dt) = \int_0^\infty [1 - \mu]^{2C_L s} \mathbb{P}\left(\frac{t_0}{C_L} \in ds\right) \\ &\sim (1 - \gamma) + \gamma \int_0^\infty e^{-2C_L \log(\frac{1}{1-\mu})} e^{-s} ds = \frac{1}{1 + 2C_L \log(\frac{1}{1-\mu})}. \end{aligned} \quad (3.1.37)$$

For small μ we have $\log(\frac{1}{1-\mu}) \sim \mu$, and so we arrive at

$$\mathbb{E}([1 - \mu]^{2t_0}) \sim (1 - \gamma) + \gamma \frac{1}{1 + 2C_L \mu}. \quad (3.1.38)$$

We can now insert the various choices for C_L and γ appearing in (3.1.31–3.1.34), to obtain the asymptotic behavior of h and $h(x)$ in the various cases. For instance, for $\alpha = \infty$ we find that

$$h \sim \frac{1}{1 + 4N_{\text{eff}} \mu} \quad \text{with } N_{\text{eff}} = NL^2. \quad (3.1.39)$$

This is the same as the probability for two individuals to be identical by descent in a *single colony* Wright-Fisher model with $2N_{\text{eff}}$ individuals (recall the calculation of χ in (2.2.10) in the proof of Lemma 2.2.1)). Thus, in this limit the stepping model behaves as if all individuals are thrown into a single colony and ancestors are chosen in a uniform manner from the entire population (the *panmictic* behavior alluded to at the beginning of Section 3.1.3). In this case, apparently the spatial structure of the stepping stone model is irrelevant and it exhibits so-called *mean-field behavior*. Similarly, for $\alpha \in (0, \infty)$ we find that

$$h \sim \frac{1}{1 + 4N_{\text{eff}} \mu} \quad \text{with } N_{\text{eff}} = (1 + \alpha) \frac{L^2 \log L}{4\pi\nu} = \frac{1 + \alpha}{2\alpha} NL^2. \quad (3.1.40)$$

Here, the effective population is the total population times a moderation factor $\frac{1+\alpha}{2\alpha}$. The latter shows the effect of the spatial structure.

3.2 Hierarchical models

In this section we make a different choice for the lattice \mathcal{L} labelling the colonies, namely, the so-called *hierarchical group of order M* . We will show that in the limit as $M \rightarrow \infty$ this model displays *universal behavior on large space-time scales*. This universal behavior will turn out to be the result of repeated application of a *renormalization transformation* connecting successive hierarchical levels.

3.2.1 Hierarchically interacting diffusions

The hierarchical group of order M is the set

$$\Omega_M = \left\{ x = (x_k)_{k \in \mathbb{N}} : x_k \in \{0, 1, \dots, M-1\}, \sum_{k \in \mathbb{N}} x_k < \infty \right\}. \quad (3.2.1)$$

Think of this set as the collection of all telephone numbers, built from the alphabet $\{0, 1, \dots, M-1\}$ and ending with all zeroes. Think of x as the *genetic address* of colony x :

x_1 is the house, x_2 is the street, x_3 is the town, x_4 is the province, x_5 is the country, etc.

(See Sawyer and Felsenstein [6] for the genetic background of this choice.) The restriction $\sum_{k \in \mathbb{N}} x_k < \infty$ makes Ω_M countable. With componentwise addition modulo M , Ω_M becomes a group.

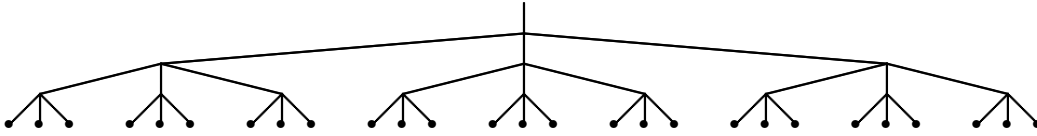


Fig. 10. The hierarchical group Ω_3 (dots) and its distance structure (lines).

On Ω_M there is a natural distance, called the *hierarchical distance*:

$$\|x - y\| = \min\{k \in \mathbb{N}_0 : x_l = y_l \ \forall l > k\}. \quad (3.2.2)$$

This is the height of the first common ancestor of x and y in the tree structure given in Fig. 10. The hierarchical distance is in fact an *ultrametric*:

$$\|x - y\| \leq \max\{\|x - z\|, \|z - y\|\} \quad \forall x, y, z \in \Omega_M. \quad (3.2.3)$$

We will be interested in the following *system of coupled stochastic differential equations*:

$$dY_x(t) = c \sum_{y \in \Omega_M} p(x, y) [Y_y(t) - Y_x(t)] dt + \sqrt{g(Y_x(t))} dW_x(t), \quad x \in \Omega_M, t \geq 0, \quad (3.2.4)$$

where

- (1) $c > 0$ is a constant;
- (2) $p(x, y)$ is a random walk transition kernel on Ω_M ;
- (3) $g: [0, 1] \rightarrow [0, \infty)$ is a diffusion function;
- (4) $\{W_x\}_{x \in \Omega_M}$ is an i.i.d. collection of standard Brownian motions.

The system in (3.2.4) models a collection of interacting WF-diffusions with migration rate c , migration kernel $p(x, y)$, and diffusion function g (recall (2.1.30)). As initial condition we take

$$Y_x(0) = \theta \in (0, 1) \quad \forall x \in \Omega_M. \quad (3.2.5)$$

For each $x \in \Omega_M$, $Y_x(t)$ denotes the fraction of individuals of type A (rather than type a) in colony x at time t (in the scaling limit given by (2.1.12)–2.1.14)). The colonies evolve via *resampling*, given by the second term in (3.2.4), and interact via *migration*, given by the first term in (3.2.4). In order for the second term to make sense, we need to place some restrictions on g . In what follows we will assume that:

- (i) $g(0) = g(1) = 0$;
- (ii) $g(u) > 0 \forall u \in (0, 1)$;
- (iii) g is Lipschitz on $[0, 1]$.

Under these restrictions, (3.2.4) has a unique strong solution. The *class* of functions satisfying (3.2.6) will be called \mathcal{H} .

In what follows we will make a *special choice for the migration kernel*, namely,

$$p(x, y) = \frac{1}{N_M} \sum_{k \geq \|x-y\|} \frac{1}{M^{2k-1}}, \quad (3.2.7)$$

where N_M is the normalizing constant. This choice, which will turn out to be particularly *well adapted to the hierarchical structure* of Ω_M , amounts to a migration mechanism in which an individual, for each $k \in \mathbb{N}$, with a probability proportional to $1/M^{k-1}$ chooses “space horizon” k , randomly chooses a colony from the k -block around its own colony, and then randomly chooses an individual from that colony. An easy computation, taking into account (3.2.3), gives $N_M = M^2/(M^2 - 1)$. This constant can be absorbed into the migration rate c .

Another motivation behind (3.2.7) is that the random walk on Ω_M with transition kernel (3.2.7) is *critically recurrent*, i.e., recurrent but barely, which is similar to the nearest-neighbor random walk on \mathbb{Z}^2 treated in Section 3.1. This property allows for rich behavior.

3.2.2 Hierarchy of space-time scales

We will look at the system along an *increasing sequence of space-time scales*. To that end, we define

$$Y_x^{[k]}(t) = \frac{1}{M^k} \sum_{y \in \Omega_M: \|y-x\| \leq k} Y_y(M^k t), \quad x \in \Omega_M, k \in \mathbb{N}. \quad (3.2.8)$$

This is the average of the components taken over a block of radius centered at x , with time speeded up proportional to the volume of the block: $M^k = \{y \in \Omega_M : \|y-x\| \leq k\}$. We will refer to $Y_x^{[k]}(\cdot)$ as the *block average around x on space-time scale k* . These block averages themselves satisfy a system of stochastic differential equations:

$$dY_x^{[k]}(t) = c \sum_{l \geq 1} \frac{1}{M^{l-1}} [Y_x^{[k+l]}(M^{-l}t) - Y_x^{[k]}(t)] dt + \sqrt{\frac{1}{M^k} \sum_{\substack{y \in \Omega_M : \\ \|y-x\| \leq k}} g(Y_y(M^k t))} dW_x^{[k]}(t), \quad x \in \Omega_M, t \geq 0, k \in \mathbb{N}, \quad (3.2.9)$$

where $\{W_x^{[k]}\}_{x \in \Omega}$ are i.i.d. standard Brownian motions. This system is deduced from (3.2.4) by summing over the k -block around x , dividing by M^k , speeding up time by M^k , and using the ultrametric property of the hierarchical distance together with the special choice of the migration kernel in (3.2.7). We also use the scalings $dW_x(M^k t) \stackrel{\text{dis}}{=} \sqrt{M^k} dW_x(t)$ and $\sqrt{a}W + \sqrt{b}W' \stackrel{\text{dis}}{=} \sqrt{a+b}W''$ to move all the diffusions terms under the square root. The initial condition in (3.2.5) becomes

$$Y_x^{[k]}(0) = \theta \in (0, 1) \quad \forall x \in \Omega_M. \quad (3.2.10)$$

Looking at (3.2.9), we may get discouraged because this system seems rather more complicated than (3.2.4). However, in the limit as $M \rightarrow \infty$ a major simplification sets in, as we will now explain.

3.2.3 Local mean-field limit

$k = 0$: In this case (3.2.9) reads

$$dY_x(t) = c \sum_{l \geq 1} \frac{1}{M^{l-1}} [Y_x^{[l]}(M^{-l}t) - Y_x(t)] dt + \sqrt{g(Y_x(t))} dW_x(t). \quad (3.2.11)$$

As $M \rightarrow \infty$, only the term with $l = 1$ survives. Moreover, $Y_x^{[1]}(M^{-1}t) \rightarrow Y_x^{[1]}(0) = \theta$ for all $t \geq 0$, by (3.2.10). Therefore we obtain that $Y_x(t) \rightarrow Z_x(t)$ with $Z_x(t)$ the solution of

$$dZ(t) = c[\theta - Z(t)] dt + \sqrt{g(Z(t))} dW(t). \quad (3.2.12)$$

In other words, in the limit as $M \rightarrow \infty$ the components *decouple* and each component performs an *autonomous diffusion* with drift towards θ and with diffusion function g .

$k = 1$: As $M \rightarrow \infty$, again only the term with $l = 1$ in (3.2.9) survives. Moreover, $Y_x^{[2]}(M^{-1}t) \rightarrow Y_x^{[2]}(0) = \theta$ for all $t \geq 0$, by (3.2.10). Furthermore, for fixed t the family

$$\{Y_y(Mt)\}_{\substack{y \in \Omega_M : \\ \|y-x\| \leq 1}} \quad (3.2.13)$$

decouples and each member *converges almost instantly to the equilibrium distribution associated with* (3.2.12), with the drift towards θ replaced by a drift towards $Y_x^{[1]}(t)$, the

instantaneous value of the 1-block average. Thus, as $M \rightarrow \infty$,

$$\frac{1}{M} \sum_{\substack{y \in \Omega_M: \\ \|y-x\| \leq 1}} g(Y_y(Mt)) \rightarrow (Fg)(Y^{[1]}(t)), \quad (3.2.14)$$

where

$$(Fg)(y) = \int_0^1 g(x) \nu_{c,y,g}(dx) \quad (3.2.15)$$

with $\nu_{c,y,g}$ the equilibrium distribution associated with the diffusion

$$dZ(t) = c[y - Z(t)] dt + \sqrt{g(Z(t))} dW(t). \quad (3.2.16)$$

Therefore we conclude that $Y_x^{[1]}(t) \rightarrow Z(t)$ with $Z(t)$ the solution of

$$dZ(t) = c[\theta - Z(t)] dt + \sqrt{(Fg)(Z(t))} dW(t). \quad (3.2.17)$$

Note that this is again an *autonomous diffusion*, but this time with a different diffusion function than in (3.2.12), namely, Fg instead of g . Later we will give an explicit formula for F .

$k \geq 2$: By moving up further in the hierarchy we find, iterating the above argument, that $Y_x^{[k]}(t) \rightarrow Z(t)$ with $Z(t)$ the solution of

$$dZ(t) = c[\theta - Z(t)] dt + \sqrt{(F^k g)(Z(t))} dW(t) \quad (3.2.18)$$

with diffusion function $F^k g$, the k -th iterate of F applied to g .

Theorem 3.2.1 *The above convergence holds true for all $c > 0$, $\theta \in (0, 1)$ and $g \in \mathcal{H}$, the class defined by (3.2.6).*

Proof. The proof is beyond the scope of this course. We refer to Dawson and Greven [3] ■

The limit $M \rightarrow \infty$ is referred to as the *local mean-field limit*. In this limit each colony has a large number of neighbors with which it interacts equally strongly. For each k , the k -block feels a drift towards the value of $(k+1)$ -block around it and equilibrates fast w.r.t. the slower motion of that block. The diffusion function of the $(k+1)$ -st block is the average of the diffusion function of the constituent k -blocks w.r.t. that equilibrium.

3.2.4 Renormalization transformation

We next move on to studying the renormalization transformation F in detail. To simplify the calculations somewhat, we henceforth pick $c = 1$.

Theorem 3.2.2 For every $v \in (0, 1)$ and $g \in \mathcal{H}$, the stochastic differential equation in (3.2.16) has a unique ergodic equilibrium $\nu_{v,g}$ given by the formula

$$\nu_{v,g}(du) = \frac{1}{N_{v,g}} \left\{ \frac{1}{g(u)} \exp \left[- \int_u^v \frac{w-v}{g(w)} dw \right] \right\} du, \quad u \in (0, 1), \quad (3.2.19)$$

with $N_{v,g}$ the normalizing constant.

Proof. The equilibrium $\nu_{v,g}$ is defined by the requirement that

$$\int_0^1 (\mathcal{L}_{v,g}f)(u) \nu_{v,g}(du) = 0 \quad \forall \mathcal{D}(\mathcal{L}_{v,g}), \quad (3.2.20)$$

with

$$\mathcal{L}_{v,g} = (v-u) \frac{\partial}{\partial u} + g(u) \frac{\partial^2}{\partial u^2} \quad (3.2.21)$$

the generator of the diffusion in (3.2.16) and $\mathcal{D}(\mathcal{L}_{v,g})$ its domain. The latter domain is dense in the set of continuous functions on $(0, 1)$. Let $\mu_{v,g}$ be the term between braces in (3.2.19). Then, by partially integrating (3.2.20) after ignoring the normalizing constant, we get

$$\int_0^1 f(u) \left[- \frac{\partial}{\partial u} \{ (v-u) \mu_{v,g}(u) \} + \frac{\partial^2}{\partial u^2} \{ g(u) \mu_{v,g}(u) \} \right] du = 0 \quad \forall \mathcal{D}(\mathcal{L}_{v,g}). \quad (3.2.22)$$

Here, the boundary terms vanish because $\mu_{v,g}$ puts no mass at 0 and 1. We conclude that the term between square brackets must be zero. Removing one differentiation, we end up with

$$-(v-u) \mu_{v,g}(u) + \frac{\partial}{\partial u} \{ g(u) \mu_{v,g}(u) \} = C_{v,g} \quad (3.2.23)$$

with $C_{v,g}$ some integration constant. Solving (3.2.23), we find

$$\mu_{v,g}(u) = \frac{1}{g(u)} \exp \left[- \int_u^v \frac{w-v}{g(w)} dw \right], \quad (3.2.24)$$

where we may absorb $C_{v,g}$ into the normalizing constant. For the fine details of the above calculation, in particular, issues of integrability near the boundaries of $[0, 1]$, we refer to Baillon et al. [1]. ■

In view of Theorem 3.2.2, F takes on the form

$$(Fg)(v) = \frac{\int_0^1 g(u) \mu_{v,g}(u) du}{\int_0^1 \mu_{v,g}(u) du}, \quad v \in (0, 1). \quad (3.2.25)$$

It can further be shown that, for all $g \in \mathcal{H}$,

$$\begin{aligned} \nu_{v,g} &\implies \delta_0 \text{ as } v \downarrow 0, \\ \nu_{v,g} &\implies \delta_1 \text{ as } v \uparrow 1, \end{aligned} \quad (3.2.26)$$

so that

$$(Fg)(0) = g(0) \text{ and } (Fg)(1) = g(1). \quad (3.2.27)$$

Note that F is a *non-linear integral transform* acting on the class \mathcal{H} .

Now that we have an explicit form for F , we can study its iterations..

Theorem 3.2.3 $F\mathcal{H} \subset \mathcal{H}$.

Proof. We must show that if $g \in \mathcal{H}$ then also $Fg \in \mathcal{H}$. It is clear from (3.2.27) that Fg satisfies condition (i) in (3.2.6). From (3.2.19) we see that, for all $v \in (0, 1)$, $\nu_{v,g}$ has full mass in $(0, 1)$. Therefore Fg satisfies condition (ii) in (3.2.6). Finally, it can be shown (see Baillon et al. [1]) that F lowers the Lipschitz constant, i.e., if

$$L[g] \stackrel{\text{def}}{=} \sup_{\substack{a,b \in [0,1]: \\ a \neq b}} \left| \frac{g(a) - g(b)}{a - b} \right|, \quad (3.2.28)$$

then

$$L[Fg] \leq L[g] \quad \forall g \in \mathcal{H}. \quad (3.2.29)$$

Hence, also condition (iii) in (3.2.6) carries over. \blacksquare

According to Theorem 3.2.3, we can iterate F indefinitely and study its *orbit*

$$\{F^k g\}_{k \in \mathbb{N}_0}. \quad (3.2.30)$$

Our key result is the following *universal scaling behavior*:

Theorem 3.2.4 For all $g \in \mathcal{H}$,

$$\lim_{k \rightarrow \infty} kF^k g = g^* \quad \text{uniformly on } [0, 1], \quad (3.2.31)$$

with

$$g^*(u) = u(1 - u) \quad (3.2.32)$$

the *Wright-Fisher diffusion function*.

Proof. The proof uses the iterates of an associated sequence of Markov kernels. Fix $g \in \mathcal{H}$. Let

$$\begin{aligned} K_g(v, du) &= \nu_{v,g}(du), \\ K_g^{[k]} &= K_{F^{k-1}g} \circ K_{F^{k-2}g} \circ \cdots \circ K_g, \quad k \in \mathbb{N}. \end{aligned} \quad (3.2.33)$$

Our equilibrium $\nu_{v,g}$ satisfies four relations:

$$\begin{aligned} (a) \quad & \int_0^1 \nu_{v,g}(du) = 1, \\ (b) \quad & \int_0^1 u \nu_{v,g}(du) = v, \\ (c) \quad & \int_0^1 u^2 \nu_{v,g}(du) = v^2 + (Fg)(v), \\ (d) \quad & \int_0^1 g(u) \nu_{v,g}(du) = (Fg)(v). \end{aligned} \quad (3.2.34)$$

Relations (a) and (d) are trivial. Relations (b) and (c) are easily deduced from (3.2.19). Iteration of (3.2.34) leads to four relations for the kernel $K_g^{[k]}$:

$$\begin{aligned} (a') \quad & \int_0^1 K_g^{[k]}(v, du) = 1, \\ (b') \quad & \int_0^1 u K_g^{[k]}(v, du) = v, \\ (c') \quad & \int_0^1 u^2 K_g^{[k]}(v, du) = v^2 + k(F^k g)(v), \\ (d') \quad & \int_0^1 g(u) K_g^{[k]}(v, du) = (F^k g)(v). \end{aligned} \quad (3.2.35)$$

Now comes a clever trick. Subtract (c') from (d'), to get

$$0 \leq \int_0^1 u(1-u)K_g^{[k]}(v, du) = v(1-v) - k(F^k g)(v). \quad (3.2.36)$$

From this we see that $kF^k g \leq g^*$ uniformly on $[0, 1]$ for all $k \in \mathbb{N}_0$. This is already half of (3.2.31). To get the other half, we proceed as follows. It follows from (3.2.36) that $F^k g \rightarrow 0$ as $k \rightarrow \infty$ uniformly on $[0, 1]$. Hence, (b') and (d') give that

$$K_g^{[k]}(\cdot) \implies (1-v)\delta_0(\cdot) + v\delta_1(\cdot) \quad \text{as } k \rightarrow \infty \quad \forall v \in (0, 1), \quad (3.2.37)$$

where we use that g vanishes only at 0 and 1. Inserting the above identity into (3.2.36), we arrive at

$$v(1-v) - k(F^k g)(v) \rightarrow 0 \quad \text{as } k \rightarrow \infty \quad \forall v \in (0, 1). \quad (3.2.38)$$

Thus we have proved (3.2.31) with pointwise convergence. To get uniform convergence, we use that the family $\{kF^k g\}_{k \in \mathbb{N}_0}$ is uniformly equicontinuous on \mathcal{H} , a fact that can be deduced from the analogue of (3.2.29):

$$L[Kf] \leq L[f] \quad \forall f, g \in \mathcal{H}. \quad (3.2.39)$$

The proof of (3.2.29) is technical (see Baillon et al. [1]). ■

The beauty of Theorem 3.2.4 is that it expresses *full universality*: no matter what $g \in \mathcal{H}$ we pick as the diffusion function for the *single components* of our system (3.2.4), the diffusion function $F^k g$ for the k -blocks (defined by (3.2.8)) with k large are *close to a multiple of the Wright-Fisher diffusion function*, namely,

$$F^k g \sim \frac{1}{k}g^* \quad \text{as } k \rightarrow \infty, \quad (3.2.40)$$

In other words, Wright-Fisher is a *global attractor on large space-time scales*. In retrospect, this somewhat justifies spending an entire course on Wright-Fisher (!).

References

- [1] J.-B. Baillon, Ph. Clément, A. Greven and F. den Hollander, On the attracting orbit of a non-linear transformation arising from renormalization of hierarchically interacting diffusions, Part I: The compact case, *Can. J. Math.* 47 (1995) 3–27.
- [2] J.T. Cox and R. Durrett, The stepping stone model: New formulas expose old myths, *Ann. Appl. Prob.*
- [3] D.A. Dawson and A. Greven, Multiple scale analysis of interacting diffusions, *Probab. Theory Relat. Fields* 95 (1993) 467–508.
- [4] R. Durrett, *Probability Models for DNA Sequence Evolution*, Springer, New York, 2002.

- [5] S.N. Ethier and T.G. Kurtz, *Markov Processes; characterization and convergence*, John Wiley & Sons, New York, 1986.
- [6] S. Sawyer and J. Felsenstein, Isolation by distance in a hierarchically clustered population, *J. Appl. Probab.* 20 (1983) 1–10.

Details to look at:

- (1) Explain why Lemma 2.1.2 and (2.1.11) imply that $\mathbb{E}(\tau)$ is of order N . Since $\mathbb{E}(\tau) = \sum_{n=0}^{\infty} \mathbb{P}(\tau > n) = \sum_{n=0}^{\infty} \mathbb{P}(H_n \neq 0)$ and $H_n \leq \frac{1}{2}$, we have $\mathbb{E}(\tau) \geq 2 \sum_{n=0}^{\infty} \mathbb{E}(H_n) = 4NH_0$. How do we get an upper bound? Is there an easy formula for $\mathbb{E}(\tau)$ valid for large N ?
- (2) Exhibit the role of large N in (2.1.16).
- (3) The argument behind (2.2.11–2.2.12) is unclear, because the mutation probability depends on the type. The role of small u, v is not transparent enough.
- (4) Give an explicit computation showing why σ in (2.3.21) has distribution $\text{EXP}(1/\theta)$. Explain why Z_2 in (2.3.26) is independent of Z_1 in (2.3.23).
- (5) Provide some more background for Lemmas 3.1.4–3.1.8.
- (6) Analyze the intuitive origin of the factor $\frac{1+\alpha}{2\alpha}$ in (3.1.40).
- (7) Be more precise about the integration constant in (3.2.23).
- (8) Update reference [2].